AD_____

GRANT NO: DAMD17-94-J-4299

TITLE: The Cloning of the BRCA1 Gene

PRINCIPAL INVESTIGATOR(S): Steven A. Narod, M.D.

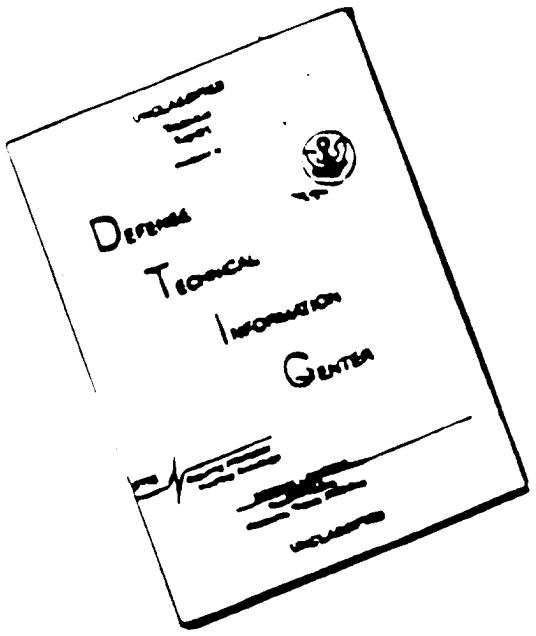
CONTRACTING ORGANIZATION: Montreal General Hospital
Research Institute
Montreal, Quebec, Canada H3G 1A4

REPORT DATE: September 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

September 1995 Annual (1 Sep 94 - 31 Aug 95)

The Cloning of the BRCA1 Gene DAMD17-94-J-4299

Steven A. Narod, M.D.

Montreal General Hospital Research Institute
Montreal, Quebec, Canada H3G 1A4

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

Approved for public release; distribution unlimited

Dr. Narod applied techniques of linkage analysis to a panel of breast-ovarian cancer families and reduced the region of assignment of the BRCA1 gene to a 600kb interval. A physical map of this region around the 17HSDB locus was created and genes in this region were identified. In collaboration with other groups, the BRCA1 gene was cloned in late 1994. The identification of this gene has led to many important questions regarding the biology of the gene, the population genetics of BRCA1 mutations and clinical management of women who are identified to be carriers of BRCA1 mutations. Dr. Narod has studied the range of BRCA1 mutations in 60 Canadian families and 20 families from the USA. Dr. Narod has established that both genetic and non-genetic factors (reproductive history) modify the risk of cancer in carriers of BRCA1 mutations. Common mutations in particular ethnic groups have been identified. Knowledge of these ethnic associations will greatly facilitate screening efforts in the high risk population.

Hereditary, Genetics, Linkage, Cloning, High-Risk, BRCA1, Humans, Anatomical Samples, Breast Cancer, Chromosome 17, Breast-Ovarian Cancer Syndrome

117

Unclassified Unclassified Unclassified Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

__✓__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature Date

Sept 7, 1995

Table of Contents

Introduction

The BRCA1 gene was mapped to chromosome 17q in 1990 by Mary- Claire King and colleagues. Shortly thereafter, Dr. Narod was able to confirm that BRCA1 was the gene for the breast-ovarian cancer syndrome as well. From 1990-1994, a series of experiments in several laboratories confirmed that BRCA1 was the gene responsible for the majority of families with multiple cases of breast and ovarian cancer. The risk of breast cancer for women who carry a mutant copy of this gene is 87% to age 70. The risk of ovarian cancer is 50%. The purpose of the present study is to clone the BRCA1 gene and to identify the range of mutations present in the families with the breast ovarian cancer syndrome.

The objectives of the study were stated as follows:

1) To identify highly informative polymorphic markers in the region of BRCA1.

2) To type linked breast and ovarian cancer families with these markers and to identify all of the crossovers which provide information about the minimum region of chromosome 17 containing BRCA1.

3) To produce a physical map of the minimum region by identifying a series of overlapping contiguous fragments of YAC DNA inserts that span the region.

4) To identify the coding sequences in the cloned YACs and to assay the expression of these genes in normal breast and ovary and in tumors derived from these tissues.

5) To test genes identified to be in the minimum region for rearrangements and for point mutations that may be associated with cancer predisposition in the families.

6) To sequence the BRCA1 gene in our panel of 100 hereditary breast and breast-ovarian cancer families and to identify the range of mutations in this gene.

7) To evaluate the association between particular mutations and cancer patterns in the panel of families

8) To sequence the BRCA1 gene in the constitutional DNA of panel of women with apparently sporadic breast and ovarian cancers.

9) To sequence the BRCA1 gene in the tumor DNA of panel of women with apparently sporadic breast and ovarian cancers.

Objectives 1 to 5 have been accomplished through a collaboration of Dr. Narod and other researchers.

Objective 6 has been reached for 60 families and should be completed by July 1, 1995

Objective 7 is in progress. Early results do not support the hypothesis that different mutations are associated with different phenotypes. Rather, it appears that the other genes will be more important in modifying genetic risk.

Objective 8 is in progress for two groups - Ashkenazi Jewish women and French Canadian women.

Objective 9 has been studied by others. It is found that mutations in BRCA1 are not frequent in sporadic breast and ovarian tumours. The great majority of mutations detected have been germline (hereditary) mutations. This objective will not be pursued further.

Details of the progress of these individual objectives follows.

Body

6.0 Overview. The goal of the current study was to identify the BRCA1 gene through positional cloning. The approach involved a combination of genetic mapping and physical mapping, leading to the identification of candidate genes from chromosome 17q. Through a collaborative effort the BRCA1 gene was cloned in October 1994 (Miki et al, 1994). Dr. Narod was a contributing partner in this collaboration.

The cloning of the BRCA1 gene took place in year one of the grant period of Dr Narod. Dr. Narod has capitalized on the early results and has gone on to use these results to investigate several aspects of the population genetics of BRCA1, including: establishing the range of BRCA1 mutations in breast-ovary families; 2) establishing mutation/haplotype correlations; 3) establishing the existence of predominant mutations in ethnic subgroups; 4) establishing the proportion of breast cancer families due to BRCA1; 5) establishing genetic and non-genetic modifiers of BRCA1 penetrance.

This body of work has contributed to the completion of 12 papers (9 published, 1 in press, 2 submitted). These papers are listed in the References below and a copy of each is provided in the Appendix. Dr. Narod is the first author on three papers, is the senior author on four other papers and is a contributing author on the five remaining papers.

6.1 Genetic mapping of BRCA1. In 1994 the region of assignment was established to be an interval of 1.2 kb surrounding the gene for estradiol dehydrognease (EDH17B). This was a key candidate gene and had been tentatively ruled out by direct sequence analysis, but not by linkage analysis. By examining a genetic recombinant in a 45-year old woman in a family with 10 cases of early-onset breast cancer and a single case of ovarian cancer it was possible to map the BRCA1 gene distal to the EDH17B locus (Tonin et al, 1994). This

became the proximal boundary for BRCA1. Through a collaboration with the University of Utah, a distal boundary for the BRCA1 interval (D17S78) was established. These two markers formed the genetic interval of assignment of BRCA1, a region of 600kb.

6.2 Physical mapping of BRCA1. At the same time as the construction of the genetic map, a physical map of cloned DNA fragments (YACS) was under construction in the laboratory of Dr Narod, in collaboration with Dr. J. Rommens of the Hospital for Sick Children Toronto for the region corresponding to the genetic map (Rommens et al, 1995). This physical map was used to identify 30 unique cDNA fragments, by hybridization of RNA libraries to immobilized YACs. These 30 clones were aligned to form ten transcription units. Several known and unknown genes were mapped by this process, including gamma-tubulin and the Ki antigen gene. None of these 30 clones was the BRCA1 gene. The construction of the physical map using the YAC library took place in the laboratory of Dr Narod, the retrieval of gene fragments took place in the laboratory of Dr. Rommens and the sequencing of the fragments took place in the laboratory of Dr. Jacques Simard.

6.3 Identification of BRCA1 mutations. The BRCA1 gene was identified in September 1994. A number of mutations were identified in one of the candidate genes by sequencing of tumor DNA and constitutional DNA, in the laboratory of Mark Skolnick and Myriad Genetics, Salt Lake City, Utah. By using the primer sequences for the 23 BRCA1 coding exons it was then possible to screen individuals for germline mutations in the BRCA1 gene. A total of 30 Canadian hereditary breast cancer families were selected from over 100 families collected by Dr. Narod for the initial panel chosen for sequence analysis. Ten of these families were sequenced in the laboratory of Myriad genetics (Utah) and 20 families were sequenced in Quebec by the team of Dr Narod, including Dr. Jacques Simard in Quebec city. The PCR amplified fragments were sequenced in both the coding and non-coding orientations. This approach guaranteed the greatest degree of mutation sensitivity, which was essential at the early stages of mutation detection. A total of 14 BRCA1 mutations were identified in the 30 families (Simard et al, 1994). This experiment

was noteworthy in that the first examples of recurrent mutations were identified in the Canadian families. In four families the mutation 185 del (AG) was found. This is now the most commonly reported BRCA1 mutation. In another four families a second mutation, 5728 ins C, was found. By creating genetic haplotypes it was evident that the common mutations were seen in the context of a common haplotype. This implies that the patients with these mutations, although previously believed to have been unrelated, have a common ancestor. A second set of Canadian families has now been screened for mutations. In this set of 30 breast cancer families an additional four mutations have been found (manuscript in preparation). Three of these mutations have been seen before. The other is a novel mutation which appears to be unique to French-Canadians. In total, 18 BRCA1 mutations have been identified in 60 Canadian families.

In additional, a further 20 breast-ovarian cancer families from Creighton University have been characterized by a combination of linkage analysis, haplotype analysis and direct sequencing in collaboration with Dr. Henry Lynch (Omaha, Nebraska) and Dr. Gilbert Lenoir (Lyon, France). These families have been ascertained by Dr. Lynch. The linkage and haplotype analysis was performed by Dr. Narod in Montreal. The sequencing for BRCA1 mutations was done by Dr. Lenoir in Lyon. There have been 16 BRCA1 mutations identified in the 20 families from Creighton University. These families are the subject of a separate report (Serova et al, submitted) which is included in the Appendix. It was noteworthy that in six of the 16 families there was a reduction in the amount of BRCA1 RNA in the lymphoblastoid cells. This implies that the presence of a mutation (most frequently a nonsense or frameshift mutation) leads to RNA instability by a mechanism that is not yet understood.

The linkage and mutation data that is generated for the Creighton families is used for a genetic counselling evaluation project that is also funded by the Department of the Army (Caryn Lerman PI). Dr. Narod provides the risk assessment data used for this study and has to date provided DNA-based risk assessments to over 250 individuals in these 20

families. The funds for the laboratory component of this study are derived from the current project *The Cloning of the BRCA1 Gene.* Dr Narod does not receive any additional funding for this project from Dr. Lerman.

6.4 Genetic epidemiology of BRCA1. After the cloning of BRCA1 much of the effort in Dr. Narod laboratory focused on studies of the range and frequency of BRCA1 mutations in the population and the use of this information for the rapid detection of mutations.

It was suggested by the study of the first 30 Canadian families that a small number of common mutations could possibly account for the majority of BRCA1 mutations in the population (Simard et al, 1994). Eight of the first 14 mutations were recurrent. In every case, a recurrent mutation was seen to be associated with a single haplotype. This observation led to the hypothesis that by genotyping a small number of markers from the region surrounding the BRCA1 gene it would be possible to predict the associated germline mutation by inspection of the segregating haplotype. This would result in a great saving of effort, as it would not be necessary to sequence the entire gene. To evaluate this possibility, Dr. Narod is directing a collaboration to establish the corresponding haplotypes for 200 breast-ovary cancer families. To date, over 150 different families have been identified by collaborators and the haplotype analysis of these is underway in the laboratory of Dr. Narod. Results are available on 80 families. Early results indicate that the majority of recurrent mutations are associated with a common haplotype. However, the association between haplotype and mutation is imperfect because of genetic recombination, mutation at the marker locus, recurrent *de novo* mutation and the presence of one frequent haplotype (approx 35%) in the general population. It is anticipated that this study will be completed in March 1996.

In the paper of Simard et al (1994) there were four families reported with the 185 del AG mutation. It was noted that all of these are of Ashkenazi Jewish origin. This finding was extended to include two additional families and it was noted that these were

Ashkenazi as well (Tonin et al, 1995). This specific association has now been seen in over twenty families worldwide. It has been found by another research group that this mutation may be present in 1% of Ashkenazi Jews and this story has been the subject of much media attention recently.

One of the critical questions is to estimate the proportion of Jewish women with breast and ovarian cancer who are carriers of the del AG mutation. To answer this question a study of 300 Jewish patients with ovarian cancer is planned. These patients will be selected from the list of all living patients with epithelial ovarian cancer at each of six hospitals in Canada and the USA. Each will be interviewed about family history and reproductive history and a blood sample will be obtained to screen for the 185 del AG mutation. A control sample of 400 Jewish women will be interviewed about family and reproductive histories (no blood samples obtained from controls). Dr. Narod has received a supplement of 37,920 dollars from the National Action Committee on Breast Cancer to perform this study.

A similar study has been started on French Canadian women. 250 French-Canadian women with ovarian cancer have been selected from three Quebec hospitals. To date, 160 women have been enrolled and blood samples are now available on these. BRCA1 mutations will be sought in the familial cases. Once recurrent mutations are found, the entire study group of 250 women will be screened for the recurrent mutations. The goal of this study is to identify the frequent mutations in this ethnic group in order to plan for a provincial based screening program for hereditary breast and ovarian cancer.

Dr. Narod was one of the founders of the Breast Cancer Linkage Consortium and is one of the principal contributors. The goals of this collaborative group are to establish the genetic basis for familial breast cancer worldwide, to estimate the proportion of families attributable to each of the major breast cancer loci, and to identify the phenotypic characteristics associated with each of the major genes. Dr. Narod analyzed and prepared a

report of data from 145 breast-ovarian cancer families on behalf of the Breast Cancer Linkage Consortium (Narod et al, 1995a). The study estimated that 76% of breast ovary cancer families were attributable to BRCA1. If the family had no cases of male breast cancer and two or more cases of ovarian cancer, the estimated linked proportion was 92%. After the BRCA1 gene was cloned and BRCA2 was mapped it was possible to reanalyze the data from the 145 families (Narod et al, 1995b). There were 10 apparently unlinked families in the original report. BRCA1 mutations were subsequently identified in three families and the other seven families are linked to BRCA2. None of the original 145 families is convincingly unlinked to BRCA1 and to BRCA2.

Not all carriers of BRCA1 mutations develop cancer, and for those who do, the age of onset varies. Some women develop breast cancer and other develop ovarian cancer. In an attempt to determine the relevant genetic and non-genetic factors which contribute to the clinical expression of the BRCA1 gene, Dr. Narod has collected DNA samples and associated clinical information on over 300 female carriers of BRCA1 mutations. By using of a historical cohort design, and a Cox proportional hazards analysis, it has been possible to study the effect of modifying genes and of reproductive factors on cancer penetrance (Narod et al, in press, Phelan et al, submitted). It was found that parity is an important risk modifier - each additional birth decreases the risk of breast cancer by 15%, but increases the risk of ovarian cancer by 40%. However, the risk of ovarian cancer is decreased by a late birth. There was a strong cohort effect present for both cancer types; the risk of breast and ovarian cancer is roughly double for women born after 1930 than for women born before 1930. Several genetic polymorphisms have been evaluated to see if they modify the penetrance of BRCA1. The presence of a rare allele of the HRAS1 polymorphism is associated with a 2.85-fold increase in the risk of ovarian cancer in the cohort of 307 BRCA1 carries ($p = 0.002$). There was no effect on the penetrance of breast cancer by the HRAS1 locus. Additional genes that are being evaluated for modifying effects on this cohort include epoxide hydrolase and GSTM1.

In this panel of 80 families there was no significant association between mutation type or mutation position and the risk of breast or ovarian cancer. The mutations in the 3' end of the gene were associated with a 40% reduction in ovarian cancer risk, but this difference was not significant ($p = 0.28$).

7. Conclusions. Through a combined linkage and physical mapping approach Dr. Narod contributed to the cloning of the BRCA1 gene in late 1994. The identification of this gene has led to many important questions regarding the biology of the gene, the population genetics of BRCA1 mutations and clinical management of women who are identified to be carriers of BRCA1 mutations. Dr. Narods efforts are now focused on developing rapid methods for detecting BRCA1 mutations in the population by taking advantage of the haplotype and ethnic associations of known mutations. Collaborative Networks are being established to enhance the study of the modifying genes which may jointly determine the range of expression and the penetrance of the gene. This information will be useful in applying molecular technology to the screening and counselling of women at high genetic risk for breast and ovarian cancer.

References pertaining to this grant

1. Miki et al, and 44 others, including Narod SA. A strong candidate gene for the breast and ovarian cancer susceptibility gene, BRCA1. Science (1994) 266, 66-71.

2. Narod SA, Ford D, Devilee P, Barkardottir RB, Lynch HT, Smith SA et al (1995). An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Amer J Hum Genet 56:254-264.

3. Narod SA, Ford D, Devilee P, Barkardottir RB, Eyfjord J, Lenoir G, Serova O, Easton D, Goldgar D (1995b) Genetic heterogeneity of breast-ovarian cancer revisited. Amer J Human Genet 57: 958-960

4. Narod SA et al. Risk modifiers in BRCA1 carriers. International Journal of Cancer (in press).

5. Neuhausen SL, Swensen J, Miki Y, Liu Q, Tavtigian S, Shattuck-Eidens D, Kamb D, Hobbs MR, Gingrich J, Shizuya H, Kim U-J, Cochran C, Futreal PA, Wiseman RW, Lynch HT, Tonin P, Narod SA, Cannon-Albright L, Skolnick MH, and Goldgar DE. (1994) A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene BRCA1. Human Molecular Genetics 3(11):1919-1926.

6. Phelan CM, Rebbeck TR, Weber BL, Devilee P, Lynch HT, Lenoir GM, Stratton MR, Easton DF, Ruttledge MH,Larsson C, Goldgar DE, Narod SA. The risk of ovarian cancer in BRCA1 carriers is modified by the HRAS1 locus. Submitted to Nature Genetics

7. Rommens JM, Durocher F, McArthur J, Tonin P, Leblanc J-F, Allen T, Samson C, Ferri L, Narod SA, Morgan K, and Simard J. (1995) Generation of a Transcription Map at the HSD17B Locus Centromeric to BRCA1 at 17q21. Genomics 28, 530-542.

8. Serova O, Narod SA, Tonin P, Feunteun J, Sylla B, Lynch HT, Lenoir GM. BRCA1 mutation analysis of 20 breast-ovarian cancer families. Submitted to American Journal of Human Genetics.

9. Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod SA, Couch F et al (1995) A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. J Amer Medic Assoc 273:535-541.

References (continued)

10. Simard J, Tonin P, Durocher F, Morgan K, Rommens J, Gingras S, Samson C, Leblanc J-F, Belanger C, Dion F, Liu Q, Skolnick M, Goldgar D, Shattuck Eidens D, Labrie F, and Narod SA. (1994) Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. Nature Genetics 8: 392-398.

11. Tonin P, Serova O, Simard J, Lenoir G, Feunteun J, Morgan K, Lynch H, Narod SA. (1994) The gene for hereditary breast-ovarian, BRCA1 maps distal to EDH17B2 in chromosome region 17q12-21. Hum Molec Genet 3(9):1679-1682.

12. Tonin P, Serova O, Lenoir G, Lynch H, Durocher F, Simard J, Morgan K, and Narod SA. (1995) BRCA1 mutations in Ashkenazi Jewish women. American Journal of Human Genetics 57:189.

Appendix

A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*

Yoshio Miki, Jeff Swensen, Donna Shattuck-Eidens, P. Andrew Futreal, Keith Harshman, Sean Tavtigian, Qingyun Liu, Charles Cochran, L. Michelle Bennett, Wei Ding, Russell Bell, Judith Rosenthal, Charles Hussey, Thanh Tran, Melody McClure, Cheryl Frye, Tom Hattier, Robert Phelps, Astrid Haugen-Strano, Harold Katcher, Kazuko Yakumo, Zahra Gholami, Daniel Shaffer, Steven Stone, Steven Bayer, Christian Wray, Robert Bogden, Priya Dayananth, John Ward, Patricia Tonin, Steven Narod, Pam K. Bristow, Frank H. Norris, Leah Helvering, Paul Morrison, Paul Rosteck, Mei Lai, J. Carl Barrett, Cathryn Lewis, Susan Neuhausen, Lisa Cannon-Albright, David Goldgar, Roger Wiseman, Alexander Kamb, Mark H. Skolnick*

A strong candidate for the 17q-linked *BRCA1* gene, which influences susceptibility to breast and ovarian cancer, has been identified by positional cloning methods. Probable predisposing mutations have been detected in five of eight kindreds presumed to segregate *BRCA1* susceptibility alleles. The mutations include an 11–base pair deletion, a 1–base pair insertion, a stop codon, a missense substitution, and an inferred regulatory mutation. The *BRCA1* gene is expressed in numerous tissues, including breast and ovary, and encodes a predicted protein of 1863 amino acids. This protein contains a zinc finger domain in its amino-terminal region, but is otherwise unrelated to previously described proteins. Identification of *BRCA1* should facilitate early diagnosis of breast and ovarian cancer susceptibility in some individuals as well as a better understanding of breast cancer biology.

Breast cancer is one of the most common and important diseases affecting women. Current estimates indicate that one in eight American women who reach age 95 will develop breast cancer (*1*). Treatment of advanced breast cancer is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer, is often rapidly fatal and is the fourth most common cause of cancer mortality in American women.

Y. Miki, J. Swensen, K. Yakumo, C. Lewis, S. Neuhausen, and D. Goldgar are in the Department of Medical Informatics, University of Utah Medical Center, Salt Lake City, UT 84132, USA. D. Shattuck-Eidens, K. Harshman, S. Tavtigian, Q. Liu, W. Ding, R. Bell, J. Rosenthal, C. Hussey, T. Tran, M. McClure, C. Frye, T. Hattier, R. Phelps, H. Katcher, Z. Gholami, D. Shaffer, S. Stone, S. Bayer, C. Wray, R. Bogden, P. Dayananth, and A. Kamb are at Myriad Genetics, 421 Wakara Way, Salt Lake City, UT 84108, USA. P. A. Futreal, C. Cochran, L. M. Bennett, A. Huagen-Strano, J. C. Barrett, and R. Wiseman are at the Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA. J. Ward and L. Cannon-Albright are in the Department of Internal Medicine, University of Utah Medical Center, Salt Lake City, UT 84132, USA. P. Tonin and S. Narod are in the Department of Medical Genetics, McGill University, Montreal, Quebec, H3G 1A4, Canada. P. K. Bristow, F. H. Norris, L. Helvering, P. Morrison, P. Rosteck, and M. Lai are at Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA. M. H. Skolnick is in the Department of Medical Informatics, University of Utah Medical Center, and Myriad Genetics, Salt Lake City, UT 84108, USA.

*To whom correspondence should be addressed.

Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 30 (*2*). Breast cancer has been subdivided into two types, early-onset and late-onset, a division that is based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, *BRCA1*, is thought to account for approximately 45% of families with significantly high breast cancer incidence and at least 80% of families with increased incidence of both early-onset breast cancer and ovarian cancer (*3*). Intense efforts to isolate the *BRCA1* gene have proceeded since it was first mapped to chromosome arm 17q in 1990 (*4*, *5*). A second locus, *BRCA2*, recently mapped to chromosome arm 13q (*6*), appears to account for a proportion of early-onset breast cancer roughly equal to that resulting from *BRCA1*. Unlike *BRCA1*, however, BRCA2 may not influence ovarian cancer risk. The remaining susceptibility to early-onset breast cancer is likely attributable to unmapped genes for familial cancer and rare germline mutations in genes such as *TP53*, which encodes the tumor suppressor protein p53 (*7*). It has also been suggested that heterozygote carriers of defective forms of the gene predisposing to ataxia telangiectasia are at higher risk for breast cancer (*8*, *9*). Late-onset breast cancer is often familial in origin, although the risks in relatives are not as high as those for early-onset breast cancer (*10*, *11*). The percentage of such cases that are due to genetic susceptibility is unknown.

Like many other genes involved in familial cancer, *BRCA1* appears to encode a tumor suppressor, a protein that acts as a negative regulator of tumor growth. Cancer-predisposing alleles typically carry mutations that cause loss or reduction of gene function. Predisposition to cancer is inherited as a dominant genetic trait, whereas the predisposing allele generally behaves as a recessive allele in somatic cells. Thus, a single inherited copy of the mutant allele causes predisposition, and loss or inactivation of the wild-type allele completes one of the steps in progression toward malignancy. When chromosome loss is observed in breast and ovarian tumors from patients who carry *BRCA1* predisposing alleles, the wild-type copy of *BRCA1* is invariably lost while the presumptive mutant allele is retained (*12–14*). This finding supports the hypothesis that *BRCA1* is a tumor suppressor gene and suggests that the functional BRCA1 protein is present in normal breast and ovarian epithelium tissue and is altered, reduced, or absent in some breast and ovarian tumors.

Genetic analysis of recombinant chromosomes in members of large kindreds allowed localization of *BRCA1* initially to a region of 1 to 2 megabases on chromosome 17q (*15–17*) and, subsequently, to a region of about 600 kilobase pairs (kb) (*18*) between markers *D17S1321* and *D17S1325* (*19*). A physical map comprised of overlapping yeast artificial chromosomes (YACs), P1, bacterial artificial chromosomes (BACs), and cosmid clones was generated for this region (*18*).

Identification of a strong BRCA1 candidate gene. Several strategies were used to develop a detailed map of transcripts for the 600-kb region of 17q21 between *D17S1321* and *D17S1325*. Sixty-five candidate expressed sequences (*20*) within this region were identified. Expressed sequences were characterized by DNA sequence, database comparison, transcript size, expression pattern, genomic structure and, most importantly, DNA sequence analysis in individuals from kindreds that segregate 17q-linked breast and ovarian cancer susceptibility. Three expressed sequences eventually were merged into a single transcription unit whose characteristics strongly suggest that it is *BRCA1* (*21*). This transcription unit is located in the center of the 600-kb region (Fig. 1) spanning *D17S855* and will be referred to herein as *BRCA1*.

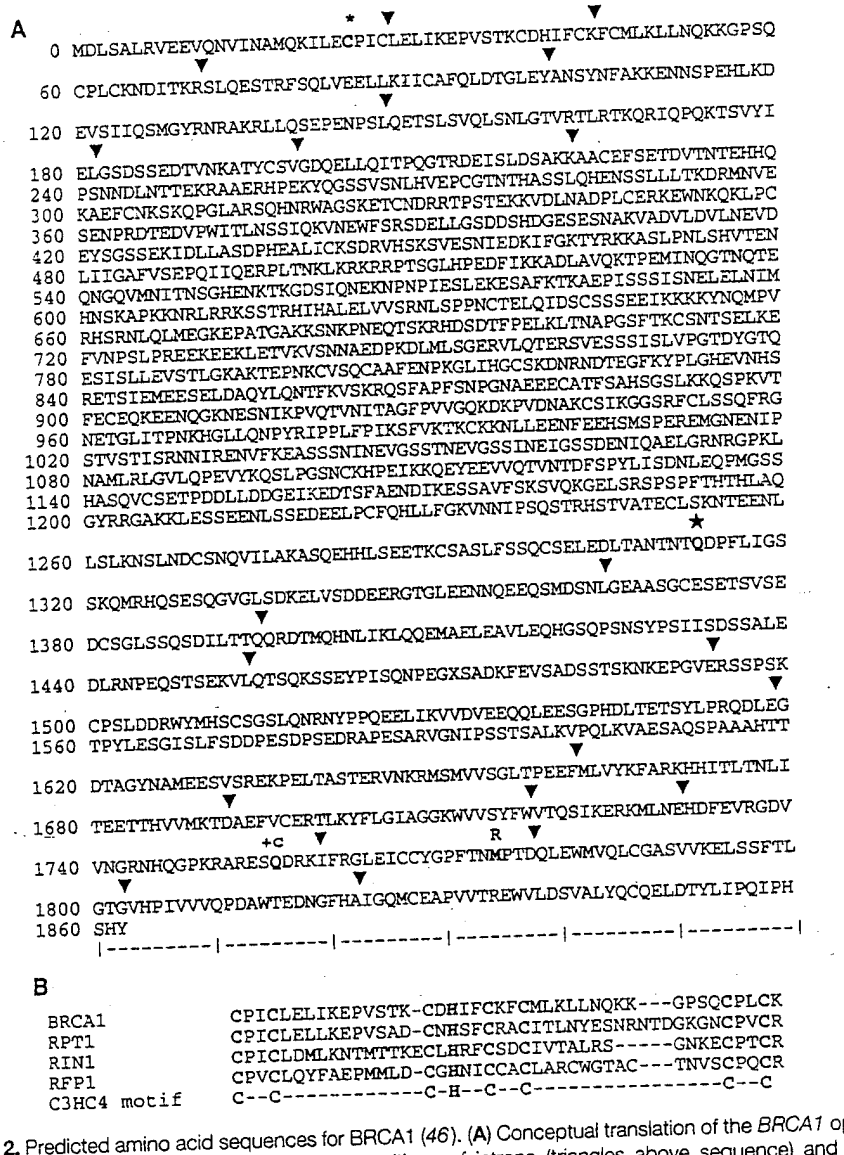
A combination of sequences obtained from complementary DNA (cDNA) clones, hybrid-selected sequences, and amplified polymerase chain reaction (PCR) products

allowed the construction of a composite, full-length *BRCA1* cDNA. The cDNA clone extending farthest in the 3′ direction contains a polyadenylate tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single, long open reading frame with a presumptive initiation codon flanked by sequences resembling the Kozak consensus sequence (*22*). This reading frame encodes a protein of 1863 amino acids (Fig. 2A). Smith-Waterman (*23*) and BLAST (*24*) searches identified a sequence near the NH_2-terminus that has considerable similarity to zinc finger domains (*25*) (Fig. 2B). This sequence contains cystine and histidine residues present in the consensus Cys_3-His-Cys_4 (C3HC4) zinc finger motif and shares many other residues with zinc finger proteins in the databases. The *BRCA1* gene is composed of 22 coding exons distributed over roughly 100 kb of genomic DNA (Fig. 3).

Hybridization of RNA blots to labeled fragments of *BRCA1* cDNA revealed a single transcript of 7.8 kb. This transcript is most abundant in testis and thymus, but is also present in breast and ovary (Fig. 4). The cDNA clones derived from the 5′ one-third of *BRCA1* transcripts display a complex pattern of alternative splicing. Four alternative splices were observed downstream of the start codon as independent cDNA clones (P3, P4, B31, and B21 in Fig. 3); three of these patterns were detected in breast cDNA (P3, B31, and B21) and two in ovary cDNA (P3 and B21). In addition, PCR analysis of cDNA samples prepared from breast, ovary, testis, and lymphocyte messenger RNA (mRNA) indicates that there is considerable heterogeneity in splice junction usage near the 5′ end of *BRCA1* transcripts, upstream of the presumptive initiation codon. How this alternative splicing is coordinated with alternative splicing farther downstream, and whether all the splice variants produce proteins with an identical NH_2-terminus, are questions that remain to be explored.

We also probed genomic DNA samples from several different species with *BRCA1* sequences devoid of the zinc finger region. Low-stringency blots revealed strongly hybridizing fragments in tissues from humans, mice, rats, rabbits, sheep, and pigs, but not chickens (Fig. 5). These results suggest that *BRCA1* is conserved in mammals.

Germline BRCA1 mutations in 17q-linked kindreds. Identification of a candidate gene as *BRCA1* requires a demonstration of potentially disruptive mutations in that gene in carrier individuals from kindreds that segregate 17q-linked susceptibility to breast and ovarian cancer. Such individuals must contain *BRCA1* alleles that differ from the wild-type sequence. The set of DNA samples used in this analysis consisted of DNA from individuals represent-



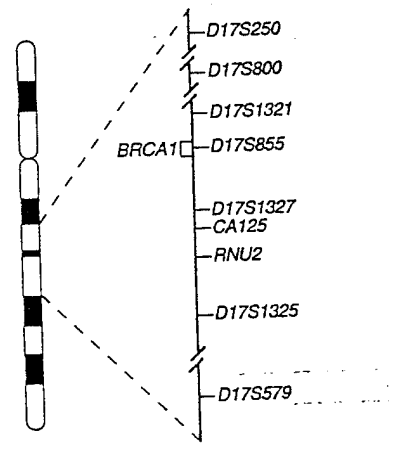


Fig. 1. Schematic map of human chromosome 17. The pertinent region containing *BRCA1* is expanded to indicate the relative positions of two previously identified genes, *CA125* (*34*) and *RNU2* (*45*). *D17S855* is located within *BRCA1*.

Fig. 2. Predicted amino acid sequences for BRCA1 (*46*). (**A**) Conceptual translation of the *BRCA1* open reading frame, indicating the approximate positions of introns (triangles above sequence) and the locations of germline mutations (boldface residues). The −11 bp deletion in kindred 1901 is shown by an asterisk; the nonsense mutation in kindred 2082 is shown by a star; the frameshift in kindred 1910 is shown by "+c"; and the missense mutation in kindred 2099 is shown by "R". The BRCA1 nucleotide sequence is deposited in GenBank with accession number U14680. PCR primer sequences are available via anonymous FTP at the following internet address: morgan.med.utah.edu in the directory pub/BRCA1; or by fax at the following number: 801-584-3650. (**B**) Alignment of the *BRCA1* zinc finger domain with three other zinc finger domains that scored highest in a Smith-Waterman alignment. RPT1 is a protein that appears to be a negative regulator of the interleukin-2 receptor in mice (*47*). RIN1 is a DNA binding protein that includes a RING finger motif related to the zinc finger (*48*). RFP1 is a putative transcription factor comprising the NH_2-terminal domain of the *RET* oncogene product (*49*). The C3HC4 motif shows the positions of the cystines and the histidine that form the zinc binding pockets.

ing eight different *BRCA1* kindreds (Table 1). The lod scores (likelihood ratios for linkage) in these kindreds range from 9.49 to −0.44 for a set of markers in 17q21. Four of the families have convincing lod scores for linkage, and four have low positive or negative lod scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least three affected members. Furthermore, all kindreds in the set display early-onset breast cancer, and four of the kindreds include at least one case of ovarian cancer, both hallmarks of *BRCA1* kindreds. Kindred 2082 has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population (*17*). All but two of the kindreds were ascertained in Utah. Kindred 2035 is from the midwestern United States. Kindred 2099 is an African American kindred from the southern United States; all other kindreds are Caucasian.

In the initial screen for predisposing mutations in *BRCA1*, DNA from one individual carrying the predisposing haplotype from each kindred was tested. The 21 coding exons and associated splice junctions were amplified from either genomic DNA samples or cDNA prepared from lymphocyte mRNA (*26*). When the amplified DNA sequences were compared to the wild-type sequence, four of the eight kindred samples were found to contain sequence variants (Table 2). All four sequence variants are heterozygous, and each appears in only one of the kindreds. Kindred 1901 contains an 11–base pair (bp) deletion in exon 2 (Cys24 frameshift to 36 Stop). Kindred 2082 contains a nonsense mutation in coding exon 11 (Gln1313 to Stop) (Fig. 6A). Kindred 1910 contains a single nucleotide insertion in coding exon 20 (Gln1756 frameshift to 1829 Stop) (Fig. 6B), and kindred 2099 contains a missense mutation in coding exon 21 (Met1775Arg). The frameshift and nonsense mutations are likely to disrupt the function of the BRCA1 proteins. The protein encoded by the insertion allele in kindred 1910 would contain an altered sequence beginning 107 amino acids residues from the wild-type COOH-terminus. The effect of the 11-bp deletion in kindred 1901 would be even more dramatic because it occurs at the twenty-fourth codon. This deletion removes the last 11 bp of exon 2 and begins at the first cystine of the zinc finger motif, thereby removing the zinc finger domain. The mutant allele in kindred 2082 would encode a protein missing 548 residues from the COOH-terminus. The missense mutation observed in kindred 2099 is potentially disruptive as it substitutes a large, charged amino acid (Arg) for a small, hydrophobic amino acid (Met). Five common polymorphisms were also identified in the *BRCA1* coding sequence (Table 3).

The individual studied in kindred 2035 is likely to carry a regulatory mutation in *BRCA1*. In her cDNA, two polymorphic sites (PM1 and PM7) appeared homozygous, whereas her genomic DNA revealed heterozygosity at these positions (Fig. 6C). One possible explanation for this observation is that mRNA from her mutant *BRCA1* allele is absent because of a mutation that affects RNA production or stability. We explored this possibility further by examining three additional polymorphic sites (PM6, PM7, and PM2) in the *BRCA1*

Table 1. Kindred descriptions and associated lod scores (*50*). Br, breast cancer; Br <50, breast cancer diagnosed under age 50; Ov, ovarian cancer.

Kindred	Cases (*n*)			Sporadic cases† (*n*)	Lod score	Markers
	Total Br	Br <50	Ov			
2082	31	20	22	7	9.49	*D17S1327*
2099	22	14	2*	0	2.36	*D17S800* and *D17S855*‡
2035	10	8	1*	0	2.25	*D17S1327*
1901	10	7	1*	0	1.50	*D17S855*
1925	4	3	0	0	0.55	*D17S579*
1910	5	4	0	0	0.36	*D17S579* and *D17S250*‡
1911	8	5	0	1	−0.20	*D17S250*
1927	5	4	0	1	−0.44	*D17S250*

*Kindred contains one individual who had both breast and ovarian cancer; this individual is counted as both a breast cancer case and as an ovarian cancer case. †Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the *BRCA1*-linked haplotype. ‡Both markers were used to calculate multipoint lod scores.

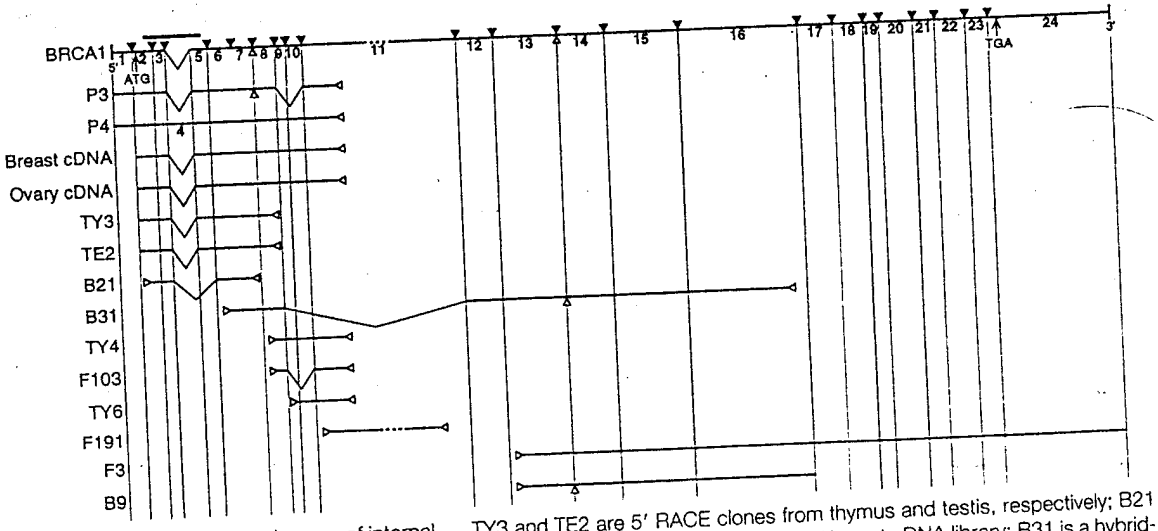


Fig. 3. Diagram of *BRCA1* mRNA, showing the locations of introns and the variants of *BRCA1* mRNA produced by alternative splicing. The top cDNA (*BRCA1*) is the composite used to generate the protein sequence in Fig. 2. Intron locations are shown by filled triangles, and the exons are numbered below the composite cDNA. Alternative mRNAs identified as cDNA clones or in hybrid-selection experiments are shown below the composite. The start codon (ATG) and stop codon (TGA) are indicated. The zinc finger region is denoted noted by a double line. "V" lines connecting exons indicate the absence of internal exons. All exons are drawn proportionally except exon 11 (indicated with a dotted line). Upward-pointing unfilled triangles show the position of a single codon (CAG) found at the start of exons 8 and 14 in some cDNAs. Leftward- and rightward-pointing unfilled triangles represent partial exons in some cDNAs. P3 and P4 are cDNA clones isolated from a placental cDNA library; TY3 and TE2 are 5′ RACE clones from thymus and testis, respectively; B21 and B9 are cDNA clones from a normal breast cDNA library; B31 is a hybrid-selected cDNA clone from breast cDNA; TY4 and TY6 are cDNA clones isolated from a thymus cDNA library; and F191, F103, and F3 are cDNA clones isolated from a fetal brain library. The *BRCA1* variants labeled breast cDNA and ovary cDNA are the major forms detected in these tissues by PCR.

coding region, which are separated by as much as 3.5 kb in the *BRCA1* transcript. In all cases where her genomic DNA appeared heterozygous for a polymorphism, her cDNA appeared homozygous. In individuals from other kindreds and in nonhaplotype carriers in kindred 2035, these polymorphic sites appeared heterozygous in cDNA, implying that amplification from cDNA was not biased in favor of one allele. This analysis indicates that a *BRCA1* mutation in kindred 2035 either prevents transcription or causes instability or aberrant splicing of the *BRCA1* transcript.

Cosegregation of BRCA1 mutations with BRCA1 haplotypes and population frequency analysis. In addition to potential disruption of protein function, a sequence variant must meet two other criteria to qualify as a candidate predisposing mutation: It must be present in members of the kindred who carry the predisposing haplotype and absent from other members of the kindred, and it must be rare in the general population.

To test for cosegregation of mutations with the corresponding *BRCA1* susceptibility allele, we screened several individuals from kindreds 1901, 1910, 2082, and 2099, including both carriers and noncarriers of the predisposing haplotype (Fig. 6). In each kindred, the corresponding mutant allele was detected only in individuals carrying the *BRCA1*-associated haplotype. In the case of the potential regulatory mutation in kindred 2035, cDNA and genomic DNA from carriers in the kindred were compared for heterozygosity at polymorphic sites. In every instance, the extinguished allele in the cDNA sample was shown to lie on the chromosome that carries the *BRCA1* predisposing allele.

To exclude the possibility that the mutations were simply common polymorphisms in the population, we used allele-specific oligonucleotides (ASOs) for each mutation to screen a set of control DNA samples (27). The actual mutation in kindred 2035 has not been identified, so we could not determine its frequency in the general population. Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African Americans were based on 39 samples used in linkage studies and on samples from 20 African Americans from Utah (28). None of the four potential predisposing mutations tested was found in the appropriate control population, indicating that they are rare in the general population (Table 2). Thus, both important requirements for *BRCA1* susceptibility alleles are fulfilled by the candidate predisposing mutations: cosegregation of the mutant allele with diseases and absence of the mutant allele in controls, indicating a low frequency in the general population (29).

Phenotypic expression of BRCA1 mutations. The effect of the mutations on the BRCA1 protein correlates with differences in the observed phenotypic expression in the *BRCA1* kindreds. Most *BRCA1* kindreds have a moderately increased ovarian cancer risk, and a smaller subset have a high risk of ovarian cancer comparable to that for breast cancer (3). Four of the five kindreds in which *BRCA1* mutations were detected fall into the former category, and the fifth (kindred 2082) falls into the group with high ovarian cancer risk. The *BRCA1* nonsense mutation found in kindred 2082 has an interesting phenotype. Kindred 2082 has a high incidence of ovarian cancer, and the mean age of breast cancer diagnosis is older than that in the other kindreds (17). This difference in age of onset could be due to an ascertainment bias in the smaller, more highly penetrant families, or it could reflect tissue-specific differences in the behavior of *BRCA1* mutations. The other four kindreds that segregate known *BRCA1* mutations have, on average, 1 ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed at an early age (late 20s or early 30s). Kindred 1910, which has a 1-bp insertion mutation, is noteworthy because three of the four affected individuals had bilateral breast cancer, and in each case the second tumor was diagnosed within a year of the first occurrence. Kindred 2035, which segregates the potential regulatory *BRCA1* mutation, might also be expected to have a

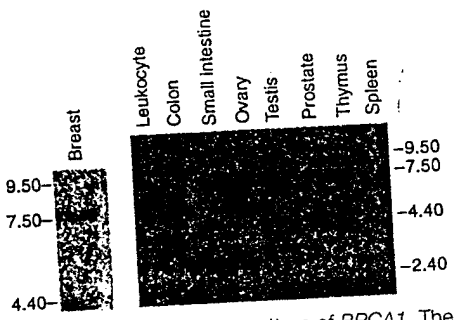


Fig. 4. Tissue expression pattern of *BRCA1*. The blots were obtained from Clontech (Palo Alto, CA) and contain RNA from the indicated tissues. Hybridization conditions were those recommended by the manufacturer, and the probe was a *BRCA1* cDNA fragment corresponding to nucleotides 3575 to 3874. Note that these tissues are heterogeneous and the percentage of relevant epithelial cells in breast and ovary can be variable. Size standards are in kilobases.

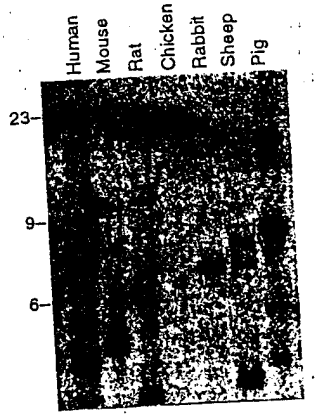


Fig. 5. Blot showing hybridization of a *BRCA1* probe to genomic DNA fragments from various species. DNA was digested with Eco RI, subjected to electrophoresis through a 0.65% agarose gel, and transferred to a nylon membrane, which was then hybridized (32) to a probe consisting of random-primed, α-^{32}P–labeled *BRCA1* cDNA sequences comprising a total of 4.6 kb. The probe excluded the zinc finger region. The final wash was at 55°C in ×2 SSPE and 1% SDS for 20 min. Size standards are in kilobases.

Table 2. Predisposing mutations in *BRCA1*. NA indicates not applicable, as the regulatory mutation is inferred and the position has not been identified.

Kindred	Codon	Mutation		Frequency in control chromosomes
		Nucleotide change	Coding effect	
1901	24	−11 bp	Frameshift or splice	0/180
2082	1313	C→T	Gln→Stop	0/170
1910	1756	Extra C	Frameshift	0/162
2099	1775	T→G	Met→Arg	0/120
2035	NA	?	Loss of transcript	NA

Table 3. Neutral polymorphisms in *BRCA1*. For the frequency in control chromosomes, the number of chromosomes with a particular base at the indicated polymorphic site is shown (A, C, G, or T).

Name	Codon location	Base in codon	Frequency in control chromosomes			
			A	C	G	T
PM1	317	2	152	0	10	0
PM6	878	2	0	55	0	100
PM7	1190	2	109	0	53	0
PM2	1443	3	0	115	0	58
PM3	1619	1	116	0	52	0

dramatic phenotype. Eighty percent of breast cancer cases in this kindred occur under age 50. This figure is as high as any in the set, suggesting that this *BRCA1* mutant allele has a high penetrance (Table 1). Kindred 1901 displays a phenotypic pattern similar to that of kindred 2035. It is likely that the 11-bp deletion beginning at codon 24 carried in kindred 1901 results in a loss of gene function similar to the effect of the regulatory mutation in kindred 2035.

Although the mutations described in this research article are clearly deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carried the mutation but lived until age 80 without developing a malignancy. It will be of utmost importance in future studies to identify other genetic factors or environmental factors that may ameliorate the effects of *BRCA1* mutations. In addition, in three of the eight putative *BRCA1*-linked kindreds, potential predisposing mutations were not found. All of these kindreds have lod scores for *BRCA1*-linked markers of less than 0.55 and thus may not truly segregate *BRCA1*-predisposing alleles. Alternatively, the mutations in these three kindreds may lie in noncoding regions of *BRCA1* and therefore have escaped detection.

The role of BRCA1 in cancer. Most mutant tumor suppressor genes identified to date encode proteins that are absent, nonfunctional, or reduced in function. The majority of *TP53* mutations are missense; some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wild-type product (*30*, *31*). A similar dominant-negative mechanism of action has been proposed for some adenomatous polyposis coli (*APC*) alleles that produce truncated molecules (*32*) and for point mutations in the Wilms tumor gene (*WT1*), which alter DNA binding of the WT1 protein (*33*). The nature of three mutations observed in the *BRCA1* coding sequence is consistent with production of either dominant-negative proteins or nonfunctional proteins. All three mutations are located in the COOH-terminal half of the protein. The regulatory mutation inferred in kindred 2035 cannot be dominant-negative; rather, this mutation likely causes reduction or complete loss of *BRCA1* expression from the affected allele. Similarly, the 11-bp deletion in kindred 1901 likely produces a nonfunctional product.

The *BRCA1* protein contains a C3HC4 zinc finger domain similar to domains found in numerous nucleic acid binding proteins. The first 180 amino acids of BRCA1 contain five more basic residues than acidic residues. In contrast, the remainder of the molecule is very acidic, with a net excess of 70 acidic residues. The excess negative charge is particularly concentrated near the COOH-terminus. Thus, one possibility is that *BRCA1* encodes a transcription factor with an NH_2-terminal DNA binding domain and a COOH-terminal "acidic blob" domain with transactivational activity. Interestingly, the product of another familial tumor suppressor gene, *WT1*, also contains zinc finger domains (*34*), and these are altered by many cancer-predisposing mutations in the gene (*33–35*). The *WT1* gene encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc finger domains alters the DNA binding properties of WT1 (*36*). Some alternatively spliced forms of *WT1* mRNA generate WT1 proteins that act as transcriptional repressors (*37*). Differential splicing of *BRCA1* may alter the zinc finger motif (Fig. 3), raising the possibility that a regulatory mechanism similar to that occurring in *WT1* may apply to *BRCA1*.

The identification of a gene that (i) falls within the interval known from genetic studies to include *BRCA1* and (ii) contains frameshift, nonsense, and regulatory mutations that cosegregate with predisposing *BRCA1* alleles strongly indicates that this gene is *BRCA1*. The observation of potential predisposing mutations in individuals

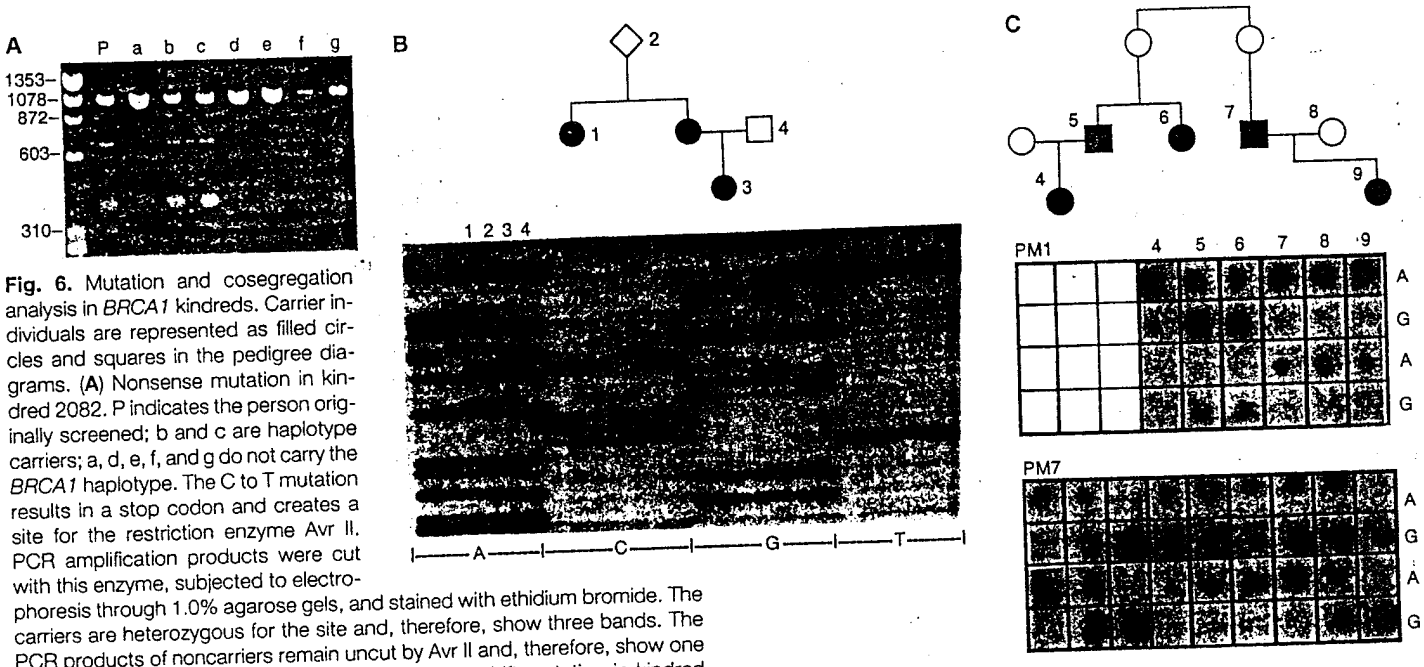


Fig. 6. Mutation and cosegregation analysis in *BRCA1* kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. (**A**) Nonsense mutation in kindred 2082. P indicates the person originally screened; b and c are haplotype carriers; a, d, e, f, and g do not carry the *BRCA1* haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme Avr II. PCR amplification products were cut with this enzyme, subjected to electrophoresis through 1.0% agarose gels, and stained with ethidium bromide. The carriers are heterozygous for the site and, therefore, show three bands. The PCR products of noncarriers remain uncut by Avr II and, therefore, show one band. Size standards are in base pairs. (**B**) Frameshift mutation in kindred 1910. Sequencing reactions were loaded side by side as A, C, G, and T reactions. The first three lanes for each nucleotide set contain sequence ladders from noncarriers. Lanes 1 to 3 contain A ladders from carrier individuals. Lane 4 contains the A ladder from a kindred member who does not carry the *BRCA1* mutation. The frameshift resulting from the cytosine insertion is apparent in lanes 1 through 3. (The diamond shape in the pedigree diagram is used to protect the confidentiality of the transmitting parent.) (**C**) Inferred regulatory mutation in kindred 2035. ASO analysis of haplotype carriers and noncarriers used two different polymorphisms, PM1 and PM7 (Table 3). Samples were examined for heterozygosity in the germ line and compared to the heterozygosity of lymphocyte mRNA. The top two rows of each panel contain PCR products amplified from genomic DNA, and the bottom two rows contain PCR products amplified from lymphocyte cDNA. "A" and "G" are the two alleles detected by the ASOs. The dark spots indicate that a particular allele is present in the sample. The first three lanes of the lower panel represent the three genotypes from unrelated individuals. There is no cDNA for sample 4 in either panel.

hose early-onset breast or ovarian cancer as not ascertained by family history supports the view that many early-onset cases e due to mutations at the *BRCA1* locus 8). The role of *BRCA1* in cancer progression may now be addressed with molecular recision. The large size and fragmented ature of the coding sequence will make xhaustive searches for new mutations chalenging. Nevertheless, the percentage of total reast and ovarian cancer caused by mutant *BRCA1* alleles will soon be estimated, and ndividual mutation frequencies and penetrances may be established. This in turn may permit accurate genetic screening for predisposition to a common, deadly disease. Although such research represents an advance n medical and biological knowledge, it also raises numerous ethical and practical issues, both scientific and social, that must be addressed by the medical community.

Note added in proof: Analysis of kindred 1911 indicates a possible linkage to *BRCA2*, suggesting that early breast cancer in this kindred is not due to a mutation of *BRCA1* (*51*).

REFERENCES AND NOTES

1. American Cancer Society, *Cancer Facts & Figures 1994* (American Cancer Society, Atlanta, GA, 1994), p. 13.
2. E. B. Claus, W. D. Thompson, N. Risch, *Am. J. Hum. Genet.* **48**, 232 (1991).
3. D. F. Easton, D. T. Bishop, D. Ford, B. P. Crockford, Breast Cancer Linkage Consortium, *ibid.* **52**, 678 (1993).
4. J. M. Hall *et al.*, *Science* **250**, 1684 (1990).
5. S. A. Narod *et al.*, *Lancet* **338**, 82 (1991).
6. R. Wooster *et al.*, *Science* **265**, 2088 (1994).
7. D. Malkin *et al.*, *ibid.* **250**, 1233 (1990).
8. M. Swift, L. Sholman, M. Perry, C. Chase, *Cancer Res.* **36**, 209 (1976).
9. M. Swift, P. J. Reitnauer, D. Morrell, C. L. Chase, *N. Engl. J. Med.* **325**, 1831 (1991).
10. L. Cannon-Albright *et al.*, *Cancer Res.* **54**, 2378 (1994).
11. C. Mettlin, I. Croghan, N. Natarajan, W. Lane, *Am. J. Epidemiol.* **131**, 973 (1990).
12. H. S. Smith *et al.*, *J. Cell. Biochem. Suppl.* **14G**, 144 (1993).
13. D. P. Kelsell, D. M. Black, D. T. Bishop, N. K. Spurr, *Hum. Mol. Genet.* **2**, 1823 (1993).
14. S. Neuhausen *et al.*, *Cancer Res.*, in press.
15. A. M. Bowcock *et al.*, *Am. J. Hum. Genet.* **52**, 718 (1993).
16. J. Simard *et al.*, *Hum. Mol. Genet.* **2**, 1193 (1993).
17. D. E. Goldgar *et al.*, *J. Natl. Cancer Inst.* **86**, 200 (1994).
18. S. L. Neuhausen *et al.*, *Hum. Mol. Genet.*, in press.
19. D. E. Goldgar *et al.*, unpublished results.
20. Candidate expressed sequences are defined as DNA sequences obtained by (i) direct screening of breast, fetal brain, lymphocyte, or ovary cDNAs (*39*) or (ii) random sequencing of genomic DNA (*40*) and prediction of coding exons by XPOUND (*41*). These expressed sequences in many cases were assembled into contigs composed of several independently identified sequences. Candidate genes may comprise more than one of these candidate expressed sequences.
21. Three independent contigs of expressed sequences 1141:1 (649 bp), 694:5 (213 bp), and 754:2 (1079 bp) were isolated by hybrid selection and eventually shown to represent portions of *BRCA1*. When expressed sequence tags (ESTs) for 1141:1 and 754:2 were used as hybridization probes for RNA blots, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, which suggested that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testis, lymphocyte, and placental cDNA libraries and PCR experiments with breast mRNA linked the 1141:1, 694:5, and 754:2 contigs. 5′ RACE experiments with thymus, testis, and breast mRNAs extended the contig to the putative 5′ end, yielding a composite full-length sequence. PCR and direct sequencing of P1s and BACs in the region were used to identify the location of introns and allowed the determination of splice donor and acceptor sites.
22. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
23. T. F. Smith and M. S. Waterman, *J. Mol. Biol.* **147**, 195 (1981).
24. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *ibid.* **215**, 195 (1990).
25. We performed database comparison by using (i) BLAST alignment algorithms (*24*) on the National Center for Biotechnology Information (NCBI) databases; (ii) Smith-Waterman alignment algorithms (*23*) on a MasPar computer to search the SwissProt database (at MasPar and at the European Molecular Biology Laboratory); and (iii) Smith-Waterman algorithms on a Compugen Biocelerator (at the Weizmann Institute) to search the GenBank (nucleotide) and SwissProt (protein) databases.
26. The templates for PCR were lymphocyte cDNA or genomic DNA from members of *BRCA1* kindreds who carried the predisposing haplotype. Sequences of PCR primers used to amplify each exon of BRCA1 are available upon request. The PCR conditions were: one cycle at 95°C (5 min); four cycles at 95°C (10 s), with the annealing temperature (T_{ann}) at 68°C for 10 s, and at 72°C for 10 s; four cycles with T_{ann} = 66°C; four cycles with T_{ann} = 64°C; four cycles with T_{ann} = 62°C; and 30 cycles with T_{ann} = 60°C. The buffer conditions were as described (*42*). The PCR products were purified from 1.0% agarose gels with Qiaex beads (QIAGEN), analyzed by cycle sequencing with [α-^{32}P]deoxy- adenosine triphosphate (*43*), and subjected to electrophoresis on 6% polyacrylamide gels. Polymorphisms were initially detected by eye on both strands and subsequently confirmed by ASO analysis (*27*).
27. PCR products were generated as described (*26*) and quantified after electrophoresis through 2% agarose gels containing ethidium bromide by comparison with DNA standards. PCR product (10 μl) was added to 110 μl of denaturant (7.5 ml of H_2O, 6.0 ml of 1 N NaOH, 1.5 ml of 0.1% bromophenol blue, and 75 ml of 0.5 mM EDTA) and incubated for 10 min at room temperature. Samples (30 μl) were then blotted onto Hybond membrane (Amersham) with a dot-blotting apparatus (Gibco-BRL). The DNA was fixed on the membrane by exposure to ultraviolet light (Stratagene). Prehybridization was carried out at 45°C in ×5 SSPE (0.75 M NaCl, 0.05 M NaH_2Po_4-H_2O, and 0.005 M EDTA) and 2% SDS (*34*). Wild-type and mutant ASOs were labeled by incubation at 37°C for 10 min in a reaction that included 5 μCi of [^{32}P]adenosine triphosphate, 100 ng of ASO, 10 units of T4 polynucleotide kinase (Boehringer-Mannheim), and kinase buffer (*44*). Labeled ASO (20 ng) was used in an overnight hybridization reaction in the same buffer used for prehybridization. Each blot was washed twice in ×5 saline sodium citrate and 0.1% SDS for 10 min at room temperature, and then for 30 min at progressively higher temperatures until nonspecific hybridization signals were eliminated. Blots were exposed to x-ray film for 40 min without an intensifying screen.
28. The African American samples from Utah were from a newborn screening program.
29. A reviewer suggested the possibility that this gene, which we call *BRCA1*, may contain frameshift mutations other than those detected here at a significant frequency in members of the general population.
30. E. Shaulian, A. Zauberman, D. Ginsberg, M. Oren, *Mol. Cell. Biol.* **12**, 5581 (1992).
31. S. Srivastava, S. Wang, Y. A. Tong, Z. M. Hao, E. H. Chang, *Cancer Res.* **53**, 4452 (1993).
32. L. K. Su *et al.*, *ibid.*, p. 2728.
33. M. H. Little *et al.*, *Hum. Mol. Genet.* **2**, 259 (1993).
34. D. A. Haber *et al.*, *Cell* **61**, 1257 (1990).
35. M. H. Little *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4791 (1992).
36. W. A. Bickmore *et al.*, *Science* **257**, 235 (1992).
37. I. A. Drummond *et al.*, *Mol. Cell. Biol.* **14**, 3800 (1994).
38. P. A. Futreal *et al.*, *Science* **266**, 120 (1994).
39. M. Lovett, J. Kere, L. M. Hinton, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9628 (1991); P. A. Futreal *et al.*, *Hum. Mol. Genet.* **3**, 1359 (1994).
40. A. Kamb *et al.*, *Nature Genet.* **8**, 22 (1994).
41. A. Thomas and M. H. Skolnick, *IMA J. Math. Appl. Med. Biol.*, in press.
42. J. Weaver-Feldhaus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7563 (1994).
43. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, ed. 2, 1989).
44. I. G. Campbell *et al.*, *Hum. Mol. Genet.* **3**, 589 (1994).
45. G. Westin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3811 (1984).
46. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
47. R. Patarca *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2733 (1988).
48. R. Lovering *et al.*, *ibid.* **90**, 2112 (1993).
49. M. Takahashi, Y. Inaguma, H. Hiai, F. Hirose, *Mol. Cell. Biol.* **8**, 1853 (1988).
50. Two of the kindreds referred to here have been studied independently [J. Feunteun *et al.*, *Am. J. Hum. Genet.* **52**, 736 (1993); D. Anderson, unpublished results].
51. D. Goldgar, unpublished results.

52. We are grateful for the cooperation of the individuals from the *BRCA1* kindreds and for the assistance of our clinic coordinators P. Fields, L. Steele, M. MacDonald, and K. Brown and for the help of C. J. Marshall. We thank D. Ballinger, K. Fournier, W. Gilbert, L. Norton, G. Omen, J. Rine, J. Simard, R. Williams, and B. Wold for scientific advice; and F. Bartholomew, H. Brownlee, S. Burgett, J. Collette, B. S. Dehoff, I. L. Jenkins, A. Leavitt, K. Richardson, and K. Rowe for technical support. Linkage study controls were kindly provided by M. Pericak-Vance. Supported in part by NIH grants CA-55914 (M.H.S.), CA-54936, CA-48711, CA-42014, CN-0522, RR-00064, and HG-00571 (D.G.), the National Cancer Institute of Canada, the Canadian Genetic Diseases Network (S.N.), and the Cedars Cancer Institute of the Royal Victoria Hospital (P.T.).

2 September 1994; accepted 14 September 1994

For information about an audio conference on the topic of the breast cancer gene *BRCA1*, see page 15.

Am. J. Hum. Genet. 56:254–264, 1995

An Evaluation of Genetic Heterogeneity in 145 Breast-Ovarian Cancer Families

Steven A. Narod,[1] Deborah Ford,[2] Peter Devilee,[3,4] Rosa B. Barkardottir,[5] Henry T. Lynch,[6] Simon A. Smith,[7] Bruce A. J. Ponder,[7] Barbara L. Weber,[8] Judy E. Garber,[9] Jill M. Birch,[10] Renee S. Cornelis,[4] David P. Kelsell,[11] Nigel K. Spurr,[11] Elizabeth Smyth,[12] Neva Haites,[13] Hagay Sobol,[14] Yves-Jean Bignon,[15] Jenny Chang-Claude,[16] Ute Hamann,[16] Annika Lindblom,[17] Ake Borg,[18] M. Steven Piver,[19] Holly H. Gallion,[7] Jeffrey P. Struewing,[20] Alice Whittemore,[21] Patricia Tonin,[1] David E. Goldgar,[22] Douglas F. Easton,[7] and the Breast Cancer Linkage Consortium

[1]Departments of Medicine and Human Genetics, Division of Medical Genetics, McGill University, Montreal; [2]Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey; Departments of [3]Pathology and [4]Human Genetics, University of Leiden, Leiden; [5]Department of Pathology, University Hospital of Iceland, Reykjavik; [6]Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha; [7]Cancer Research Campaign Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge, Cambridge; [8]Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor; [9]Dana Farber Cancer Institute, Boston; [10]Cancer Research Campaign Pediatric and Familial Cancer Research Group, University of Manchester, Manchester, United Kingdom; [11]Human Genetics Resources Laboratory, Imperial Cancer Research Fund, Potters Bar, United Kingdom; [12]Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh; [13]Department of Molecular and Cell Biology, University of Aberdeen, Aberdeen; [14]Department D'Oncologie Genetique, Institut Paoli-Calmettes, Marseille; [15]Laboratoire d' Oncologie Moleculaire, Centre Jean Perrin, Clermont-Ferrand, France; [16]Division of Epidemiology, German Cancer Research Centre, Heidelberg; [17]Department of Clinical Genetics, Karolinska Hospital, Stockholm; [18]University Hospital, Lund, Sweden; [19]Roswell Park Cancer Institute, Buffalo; [20]National Institutes of Health, Genetic Epidemiology Branch, Bethesda; [21]Division of Epidemiology, Stanford University School of Medicine, Stanford; and [22]Genetic Epidemiology Group, University of Utah, Salt Lake City

Summary

The breast-ovary cancer–family syndrome is a dominant predisposition to cancer of the breast and ovaries which has been mapped to chromosome region 17q12-q21. The majority, but not all, of breast-ovary cancer families show linkage to this susceptibility locus, designated *BRCA1*. We report here the results of a linkage analysis of 145 families with both breast and ovarian cancer. These families contain either a total of three or more cases of early-onset (before age 60 years) breast cancer or ovarian cancer. All families contained at least one case of ovarian cancer. Overall, an estimated 76% of the 145 families are linked to the BRCA1 locus. None of 13 families with cases of male breast cancer appear to be linked, but it is estimated that 92% (95% confidence interval 76%–100%) of families with no male breast cancer and with two or more ovarian cancers are linked to BRCA1. These data suggest that the breast-ovarian cancer–family syndrome is genetically heterogeneous. However, the large majority of families with early-onset breast cancer and with two or more cases of ovarian cancer are likely to be due to BRCA1 mutations.

Received December 27, 1993; accepted for publication September 30, 1994.

Address for correspondence and reprints: Dr. Steven A. Narod, Montreal General Hospital, 1650 Cedar Avenue, Room L10-107, Montreal, Quebec H3G 1A4, Canada.

0002-9297/95/5601-0033$02.00

Introduction

In 1990, a locus on chromosome 17q, BRCA1, was found to be linked to a panel of families with early-onset breast cancer (Hall et al. 1990). Shortly thereafter, this locus was shown also to predispose to cancer of the ovary (Narod et al. 1991). Not all families with hereditary breast cancer are linked to BRCA1 (Hall et al. 1990; Sobol et al. 1992); it is estimated that fewer than half of site-specific breast cancer families show linkage to BRCA1 (Easton et al. 1993). However, the proportion for families segregating ovarian cancer in addition to early-onset breast cancer is much higher. In the first report of the Breast Cancer Linkage Consortium (BCLC) it was estimated that 100% of the 57 breast-ovarian cancer families studied were linked to BRCA1, with a lower 95% confidence interval (CI) of 79% (Easton et al. 1993). Since this report, consortium members have studied many additional breast-ovarian cancer families for linkage to chromosome 17q markers, and the possibility that a number of these families may not be linked to chromosome 17q has been raised. We report here the linkage results on 145 breast-ovarian cancer families collected by the BCLC.

Families, Material, and Methods

Families

The 145 families included in this study were all those collected by the BCLC, and each contains at least three cases of either early-onset (before age 60 years) breast cancer or ovarian cancer; each family contains at least one case of ovarian cancer. All but 6 of the 57 breast-ovarian

cancer families that were included in the first report are included in this analysis, but several additional markers have now been typed, and some families have been extended to include newly identified cases. Since submission of data for the first consortium report, the other 94 families have been investigated for linkage. Nine families with site-specific ovarian cancer, as well as those families with male breast cancer, have also been reported elsewhere (Steichen-Gersdorf et al. 1994; Stratton et al. 1994).

Genetic Markers

A total of 11 markers were used to genotype the 145 families; however, the markers used varied from group to group. From centromere to telomere the markers are D17S250, THRA1, D17S800, D17S776, D17S855, ED2, D17S858, D17S579, D17S293, D17S588, and NME1. The primer sequences and map position of D17S250, THRA1, D17S579, D17S588, and D17S93 have been presented by Easton et al. (1993); D17S855 and D17858 by Anderson et al. (1993); and D17S776 and ED2 by Goldgar et al. (1994). NME1 is an RFLP from the nm23 gene region (Varesco et al. 1992).

Statistical Methods

LOD scores for linkage between BRCA1 and flanking markers were calculated using LINKAGE (Lathrop et al. 1984) and assuming (1) the disease-specific risks, in gene carriers, derived by Easton et al. (1993) and (2) a .0033 frequency of mutant BRCA1 alleles. Under this model, BRCA1 confers a breast cancer cumulative risk that is 49% by age 50 years and 71% by age 70 years and confers an ovarian cancer cumulative risk that is 16% by age 50 years and 42% by age 70 years. To implement this model in LINKAGE, different liability classes were used for unaffected females and for females affected with either breast cancer or ovarian cancer. A total of 21 liability classes were constructed, depending on either the age at cancer diagnosis or, in unaffected women, the age at last observation. The penetrance values in each liability class for unaffected females were taken from the estimated cumulative incidence figures, based on population rates for England and Wales. For affected individuals these are given by the age-specific incidence densities at the midpoint of the age group. These values have been estimated from the consortium data set by maximizing the likelihood of linkage of BRCA1 for the breast and ovarian cancer data over all possible penetrance functions. The penetrance values used are presented in table 1.These values differ slightly from those in the model used in the previous BCLC analysis, in that the new model assigns a lower probability of an older case of breast cancer being a carrier of a BRCA1 mutation and provides age-specific genotype probabilities for ovarian cancer cases.

Individuals with both breast and ovarian primary cancers were treated as if affected with breast cancer before age 30 years, in order to maximize their probability of being gene carriers. Males with breast cancer were assigned to the liability class of female cases affected before age 30 years, because of the rarity of male breast cancer.

In order to compute the multipoint LOD scores, the male map distances between markers were assumed to be 1.3 cM from D17S250 to THRA1 and 2.7 cM from THRA1 to D17S579. A 2:1 female:male genetic distance ratio was assumed. BRCA1 was placed midway between THRA1 and D17S579 (Bowcock et al. 1993; Chamberlain et al. 1993; Simard et al. 1993). Recently, BRCA1 has been assigned to a position telomeric to the markers D17S800 (Kelsell et al. 1993), D17S776 (Goldgar et al. 1994), and D17S702 (Smith et al. 1994). The relative position of BRCA1 with respect to D17S855, ED2, or D17S858 has not been determined. In most families a multipoint LOD score was computed using flanking markers D17S250 (Weber et al. 1990) and D17S579 (Hall et al. 1992). For a small number of families, data were unavailable for these two markers, and data for other markers in the region, in particular THRA1, were used.

For ease of computation, and because direct comparison of the frequencies of specific alleles across study

Table 1

Genetic Model Used in LINKAGE Analysis

Age Group (years)	Penetrance of Genotype[a]		
	dd	Dd	DD
Unaffected females:			
<30	.00025	.0237	.0237
30–39	.0024	.1357	.1357
40–49	.0109	.4222	.4222
50–59	.0266	.6732	.6732
60–69	.0468	.7946	.7946
70–79	.0700	.8586	.8586
≥80	.0960	.9000	.9000
Breast cancer cases:			
<30	.000034	.00355	.00355
30–39	.000366	.01387	.01387
40–49	.00113	.02768	.02768
50–59	.00148	.01119	.01119
60–69	.00180	.00434	.00434
70–79	.00200	.00327	.00327
≥80	.00239	.00231	.00231
Ovarian cancer cases:			
<30	.000016	.00113	.00113
30–39	.000048	.00304	.00304
40–49	.000178	.00755	.00755
50–59	.000347	.00613	.00613
60–69	.000411	.00386	.00386
70–79	.000430	.00163	.00163
≥80	.000384	.00116	.00116

[a] "D" refers to the disease allele conferring susceptibility to breast and ovarian cancers, and "d" is the normal allele. Males were assigned to the first liability class.

Table 2

Linkage Results for Breast-Ovarian Cancer Families

	LOD		
	Assuming Homogeneity	Assuming Heterogeneity	α (95% Confidence Interval)
All families (n = 145)	48.29	57.07	.76 (.62–.87)
Families with one or more cases of male breast cancer (n = 13)	−11.68	0	0 (.00–.22)
Families with no case of male breast cancer (n = 132)	59.97	62.37	.88 (.74–.97)
Families with no male case and one ovarian cancer (n = 51)	11.74	13.84	.81 (.58–.97)
Families with no male case and two or more ovarian cancers (n = 81)	48.24	48.72	.92 (.76–1.0)

groups was not possible, each of the highly polymorphic CA-repeat markers was coded as a system of equally frequent alleles (the exact number was chosen to correspond to the observed polymorphism of the marker). Haplotypes were constructed from the observed segregation of multiple closely linked markers in the extended families, assuming a minimum number of genetic recombinants.

Evidence for heterogeneity was evaluated using the admixture test, in which a proportion of families are assumed to be linked to BRCA1 and the remainder are assumed to be linked to other loci (Smith 1961). The HOMOG program (Ott 1985) was used to compute LOD scores, with heterogeneity being assumed.

Results

When the 145 families are considered together, there is clear evidence for heterogeneity of the breast-ovarian cancer syndrome (table 2). Overall, it was estimated that 76% of the 145 families studied were linked to BRCA1 (95% CI of 62%–87%). Much of the evidence for heterogeneity comes from families with one or more cases of male breast cancer; it is estimated that none of 13 families of this type are linked to BRCA1. Four of the seven families that produced multipoint LOD scores <-1.0 contain cases of male breast cancer. An example of such a family is presented in figure 1.

When the families without male cases are considered separately, the evidence for heterogeneity is much weaker. It is estimated that 88% of families without male cases are linked to BRCA1, with 95% CI of 74%–97%. Several of the families that produce negative LOD scores contain only a single case of ovarian cancer. Among the 81 breast-ovarian cancer families with two or more cases of ovarian cancer, the LOD score calculated under homogeneity was not significantly less than that calculated under heterogeneity, and the possibility that all of the families were linked to BRCA1 could not be rejected. It was also estimated that 100% of the 11 families with predominantly ovarian cancer (i.e., no male cases and no case of breast cancer under age 60 years) were linked to BRCA1 (LOD score of 2.06).

Only three families gave multipoint LOD scores <-2.0 and thereby satisfy the conventional criterion of an unlinked family (table 3); two of these families contained at least one case of male breast cancer and have been the subject of a previous report (Stratton et al. 1994). Four additional breast-ovarian cancer families give LOD scores in the range of -1.0 to -2.0. Two of these families also contained a male case. The pedigrees and haplotypes of the three families without male cases are given in figure 2, and these families are discussed individually below.

Leiden 9

This family contains nine cases of breast cancer and a single case of ovarian cancer. Three of the breast cancers occurred before the age of 40 years. Individuals 52 and 44 are sisters and were both affected with breast cancer at the age of 34 years. They have inherited different paternal chromosome 17 haplotypes, although it appears that the cancer susceptibility gene in this sibship was inherited through the father. Individual 12, also diagnosed with breast cancer at age 34 years, shares no haplotype with either of her affected cousins.

Iceland 4

In addition to eight breast cancers and one ovarian cancer, this unusual pedigree contains four cases of stomach cancer and five cases of prostate cancer. Two of the breast cancer cases are bilateral. There is no common chromosome 17 haplotype segregating among the affected members. Individual 9, who developed breast cancer at age 60 years, later developed cancer of the kidney and cancer of the ureter. Cancer of the stomach and cancer of the urinary system are seen in excess in the nonpolyposis colon cancer syndromes.

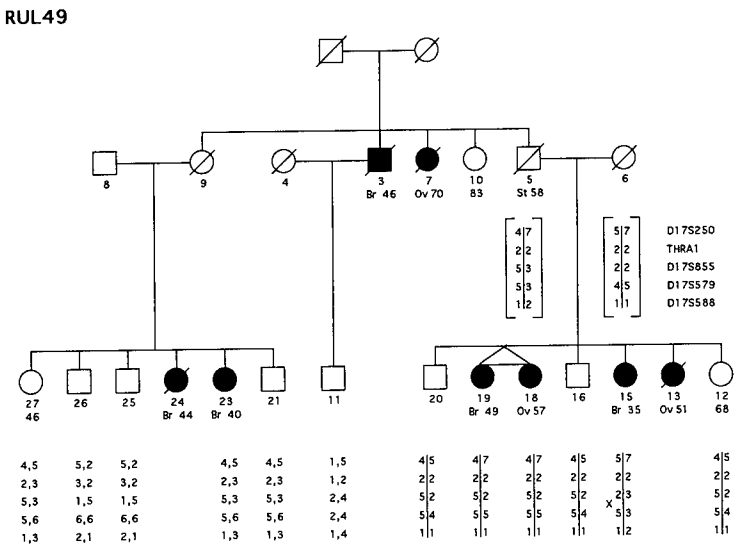


Figure 1 Pedigree of family with male breast cancer. Blackened circles indicate affected women; and a diagonal slash indicates an individual who is deceased. Individual identification numbers appear directly below the symbols. Br = breast cancer; Ov = ovarian cancer; and Sto = stomach cancer. The numbers following these abbreviations indicate age at diagnosis. The numbers arranged vertically below the individual symbols indicate the marker alleles arranged into haplotypes. Haplotypes in brackets are inferred. Marker alleles separated by a comma cannot be phased. An x indicates position of recombination.

IARC 2850

There does not seem to be a consistent haplotype segregating in this large breast-ovarian cancer family, although in the left-hand branch the (2,5,1,3) haplotype appears to segregate with the trait. This haplotype does not appear in any of the other affected women, although there have been many opportunities for genetic recombination between the different branches of the haplotype. This problem may be resolved when additional markers and individuals are tested.

In addition to these few families that are clearly unlinked to BRCA1, there were 30 families with weaker evidence against linkage (LOD scores of −0.10 to −0.99). In most of these families, a single breast cancer patient or ovarian cancer patient has a haplotype inconsistent with linkage. Some of these families could be due to other genes, but there is also the possibility that the inconsistent case could be sporadic. An example of a family of this type is CRC 128 (fig. 3), in which individual 301 was diagnosed with breast cancer at age 35 years but does not share a common haplotype with her four affected sisters, all of whom have inherited the same marker alleles, from D17S579 to NME1, from their affected mother. If this family is actually linked to BRCA1, then individual 301 appears to be a sporadic case of breast cancer. Alternatively, a double recombinant may have occurred between D17S250 and D17S579 (a distance of 8 cM in females) (LOD score of −0.91).

The cancer mutation in the left-hand branch of family IARC 2651 appears to be linked to the BRCA1 locus in all

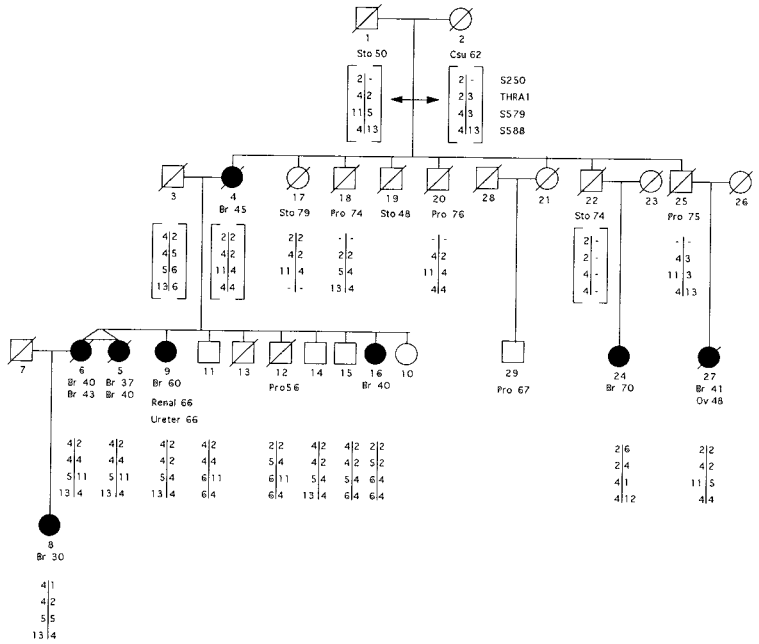
Table 3

Characteristics of Seven Breast-Ovarian Cancer Families with Multipoint LOD Scores <−1.00

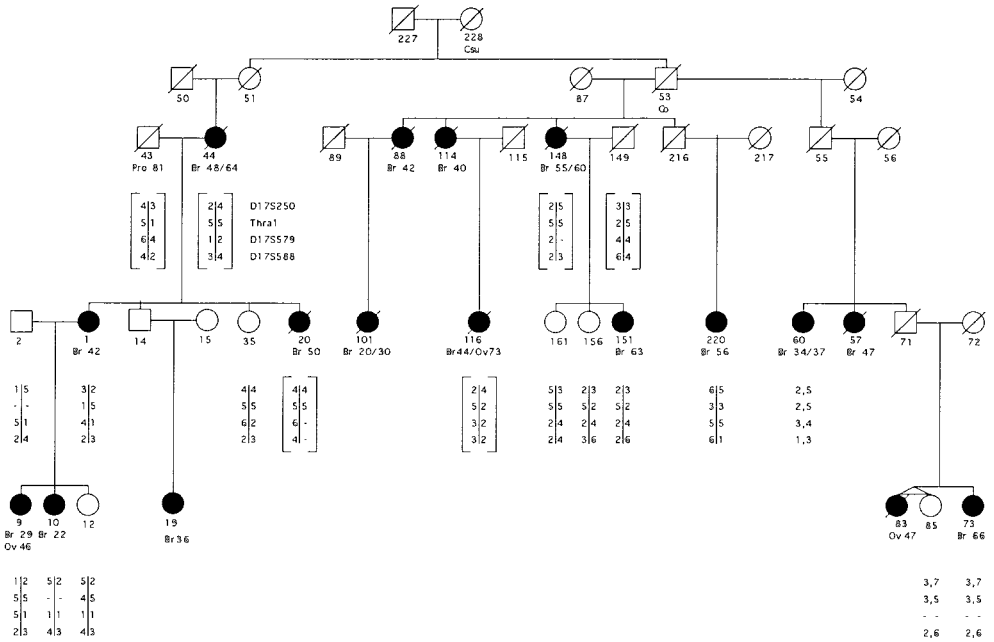
Family	No. of Breast Cancers before 60 years	No. of Male Breast Cancers	No. of Ovarian Cancers	LOD Score[a]
RUL9	7	0	1	−2.01
RUL49	4	1	3	−1.11
CRC 186	16	1	1	−2.61
IRAC 2850	15	0	2	−1.71
UTAH 2044	7	1	4	−1.32
UTAH 107	32	3	6	−2.45
ICEL 80004	6	0	1	−1.89

[a] Derived from three-point linkage analyses using the flanking markers D17S250 and D17S579.

Iceland 4



IARC 2850



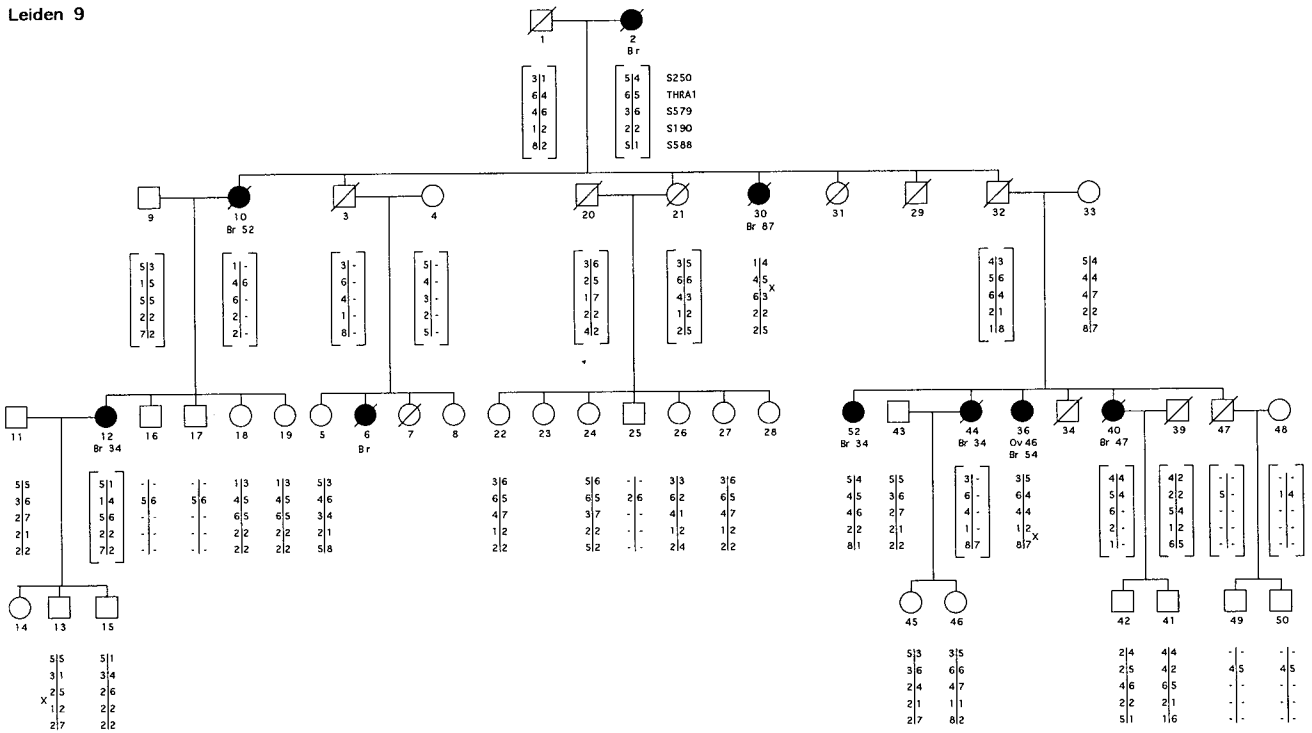


Figure 2 Pedigrees of three breast ovarian cancer families with LOD scores <−1.00. Symbols are as in fig. 1. Sto = stomach cancer; Pro = prostate cancer; and Csu = cancer, site unknown. A dash in the place of marker typing indicates missing information. The genotypes of deceased individuals in pedigree Iceland 4 are based on (noncancerous) paraffin-embedded tissue, except when inferred (*bracketed*). The horizontal arrow in family Iceland 4 means that the founder haplotypes can be inferred but not assigned.

five affected women (fig. 4). However, the linked chromosome was inherited from the grandfather, individual 69. The other three affected women in the family do not share a haplotype with the left-hand branch. Individual 120 developed a mucinous ovarian cancer at age 26 years. Under the current model, this woman would be predicted to be a BRCA1 mutation carrier. However, if the phenotype of this young woman is coded as ambiguous, then the LOD score for this family becomes 1.39. IARC 1812 is also consistent with segregation of a mutation at the BRCA1 locus, with the exception of individual 21, who is the single woman with ovarian cancer in the family (fig. 4). She developed a mucinous cystadenocarcinoma of the ovary at age 17 years and is the youngest case of ovarian cancer in the consortium data set.

Discussion

This analysis confirms our previous results, which demonstrated that a large majority of breast-ovarian cancer families are linked to BRCA1. However, in contrast to the previous analysis, under the current definition of the breast ovarian cancer family syndrome there is now clear evidence for genetic heterogeneity, with an estimated 76% of the families being linked. The negative-linkage evidence comes both from families that also contain male breast cancer cases and from families with a single case of ovarian cancer. There is little, if any, evidence for genetic heterogeneity among families with at least two cases of ovarian cancer. These data support a refined clinical definition of the breast-ovarian cancer–family syndrome—as being those families with both two or more cases of early-onset breast cancer and two or more cases of ovarian cancer. Under this refined definition, it can be expected that almost all breast-ovarian cancer families will be attributable to BRCA1 mutations.

It is important to estimate accurately the extent of genetic heterogeneity, both for purposes of genetic counseling and to expedite positional cloning. Genetic counseling using BRCA1 markers can now be considered if it can be established with reasonable certainty that a family segregates a BRCA1 mutation (Biesecker et al. 1993; King et al. 1993; Lynch et al. 1993; Narod 1993). If there is significant genetic heterogeneity in the syndrome, then it will be necessary to take this into account and to establish the probability that a given family is linked, before risk assessment is considered. In contrast, if all breast-ovarian cancer families are linked to the same locus, as the first consortium paper suggested, then counseling could be pursued in any large family of the type described here (i.e., when chance clustering is not a realistic alternative). On the basis of our current results, we feel that it is inappropriate to offer genetic counseling to breast-ovarian cancer families unless the possibility of genetic heterogeneity is taken into account. It is also reasonable at this time to consider all women who carry BRCA1 mutations to be at increased risk for both breast cancer and ovarian cancer.

Genetic linkage is currently the principal tool employed by investigators trying to identify BRCA1. This is both because there have been no reports of a chromosome 17 rearrangement segregating in a breast cancer family and because the potential for mapping BRCA1 by tumor deletions (i.e., by loss of heterozygosity) is limited. Several chromosome 17 loci may be involved in somatic mutations in breast cancers and ovarian cancers (Saito et al. 1993), and the pattern of chromosome 17 rearrangements in these tumor types is complex (Foulkes et al. 1991; Cornelis et al. 1993). Because of the importance of genetic mapping in the search for BRCA1, each informative recombinant needs to be scrutinized both in terms of the probability of linkage of the family and in terms of the number, ages, and sites of cancers in the affected women. In the pedigrees reported here, there appear to be several "sporadic" cases among families that otherwise appear to be linked to BRCA1. It is important to appreciate that crossovers that appear in affected women from presumably linked families are currently used to define the region of assignment of BRCA1 (e.g., see Simard et al. 1993; Tonin et al. 1994). If these crossovers were actually in women with sporadic cancers, or if the susceptibility gene in the family was not linked to BRCA1, then false inferences about the location of BRCA1 would result. Because of sporadic cases, and because it is now appears unlikely that all breast-ovarian cancer families are linked to BRCA1, the evidence provided by any single crossover, in regard to the mapping of BRCA1, is inconclusive.

A recent report describes a germ-line p53 mutation that segregated in a breast-ovarian cancer family (Jolly et al. 1994), and another report proposes that the proportion of these families with germ-line p53 mutations may be substantial (Buller et al. 1993). These reports suggest that p53 testing in families apparently unlinked to BRCA1 may be warranted.

However, for many of the observed pedigrees, there are explanations other than genetic heterogeneity. It is possible that some of the familial clustering is due to chance. It may be that a mutation in the BRCA1 is the cause of most, but not all, of the cases of breast cancer and ovarian cancer in a particular family. If the sporadic case is young, then there need to be several affected women in a linked family to obtain a positive LOD score. However, if the sporadic case is postmenopausal, there will be little effect on the observed LOD score, because, under the revised model, little weight is accorded to late-onset breast cancer cases (data not shown). Similarly, a family could be a hereditary site-specific breast cancer family unlinked to BRCA1 but with a single sporadic ovarian cancer case. In this case, the family would be indistinguishable from a breast-ovarian cancer family. A higher proportion of BRCA1 linkage was observed among families with two or more ovarian cancers

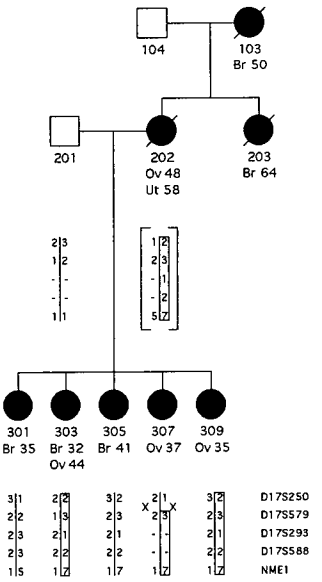


Figure 3 Pedigree CRC 128. Symbols are as in figs. 1–2. Ut = endometrial cancer. Vertical rectangles are used to indicate the most frequent haplotype found to be associated with breast or ovarian cancer in the family, but they are not intended to imply that a BRCA1 mutation is necessarily present.

(92%) than among those with a single ovarian cancer (81%). This is not unexpected, given the possibility of sporadic cancers; it is possible that our data set contains some site-specific breast cancer families with a single case of sporadic ovarian cancer (Leiden 9 and Iceland 4 are potential examples of such families). The probability of two sporadic ovarian cancers in a single family is more remote. It is also possible that on other chromosomes there exist genes that confer a high risk of breast cancer and a low risk of ovarian cancer.

Genetic recombination is an alternate explanation for discordant haplotypes, but the possibility is remote when double crossovers in a small region need be inferred. Given the density of the markers employed here to construct haplotypes, double recombination must be considered to be an unlikely possibility. However, in some families, distant pedigree branches have been typed. In the absence of genotypic information for intervening relatives, it is not possible to establish if several single crossovers have accumulated. IARC 2850 is a potential example of such a family (fig. 1). Similarly, there may be two different mutations of BRCA1 that are segregating in the same family. This possibility must also be considered in pedigrees where a number of distantly related relatives are affected but clinical information on intervening relatives is missing. Because several of the Utah families have been identified by merging cancer registry data and genealogical information, these families potentially could represent situations either where a BRCA1 mutation has entered the pedigree more than once or where a new BRCA1 mutation has occurred. Similarly, there may be unidentified consanguinity, and a single BRCA1 mutation may have entered a pedigree through two individuals thought to be unrelated.

An interesting possibility is that there is a second breast cancer susceptibility gene on chromosome 17, close to the region spanned by the markers studied here. In this case, several cases of cancer in a family will appear to be linked to BRCA1, as will one or two additional cases that appear to be sporadic but that in fact represent recombination between the BRCA1 locus and the actual susceptibility gene. The "sporadic" cases would, on average, have the same early age of diagnosis as do the other genetic cases.

Errors in diagnosis, mislabeling of specimens, or errors in assigning parentage are also possible explanations for discordant genotypes. Spurious evidence against linkage may occur if the cancer diagnosis is wrong, e.g., if the tumour site was misspecified or if the cancer was metastatic to the ovary.

It is possible that, in some cases, ductal carcinoma in situ may have been coded as breast cancer. It has not yet been established if the frequency of this precancerous condition is increased among carriers of BRCA1 mutations. This diagnosis is increasingly common because of mammographic screening of high-risk families. For example, IARC family 2090 is a linked breast-ovarian cancer family with two cases of ovarian cancer and five cases of breast cancer (fig. 5). A 42-year-old woman in this family (individual 107), who does not carry the 17q haplotype of risk, was recently diagnosed, through screening, as having ductal carcinoma in situ. Were she to be considered affected, the LOD score would diminish from 1.25 to 0.55. It will be important to determine the extent to which this condition is associated with BRCA1 mutations in these families.

It is equally important to know which histologic types of ovarian cancer are overrepresented in cancer families. In the Gilda Radner Familial Ovarian Cancer Registry, cases of mucinous adenocarcinoma and cystadenocarcinoma were underrepresented. Only 6 (1.4%) of 439 familial ovarian cancers were mucinous, in comparison with 12.7% of unselected ovarian cancers in the SEER database (Piver et al. 1993). There were no mucinous carcinomas among the 37 hereditary ovarian cancers reported by Bewtra et al. (1992), among 11 familial ovarian cancers reported by Greggi et al. (1990), or among 13 hereditary ovarian cancers in Ontario (Narod et al. 1994). In IARC families 1812 and 2651, the "unlinked" ovarian cancers were both mucinous. In contrast, only 2 of 36 ovarian cancers among BRCA1 carriers in the IARC data set were mucinous (Narod et al. 1994). Nor is it clear if borderline ovarian tumors confer familial risk. Only 5 of 439 ovarian cancers in the Gilda Radner registry were of borderline histology (Piver et al. 1993). Bewtra et al. (1992) found 2 of 37 hereditary ovarian cancers to be borderline. Data on histologic subtypes of epithelial ovarian cancer are not routinely available on all consor-

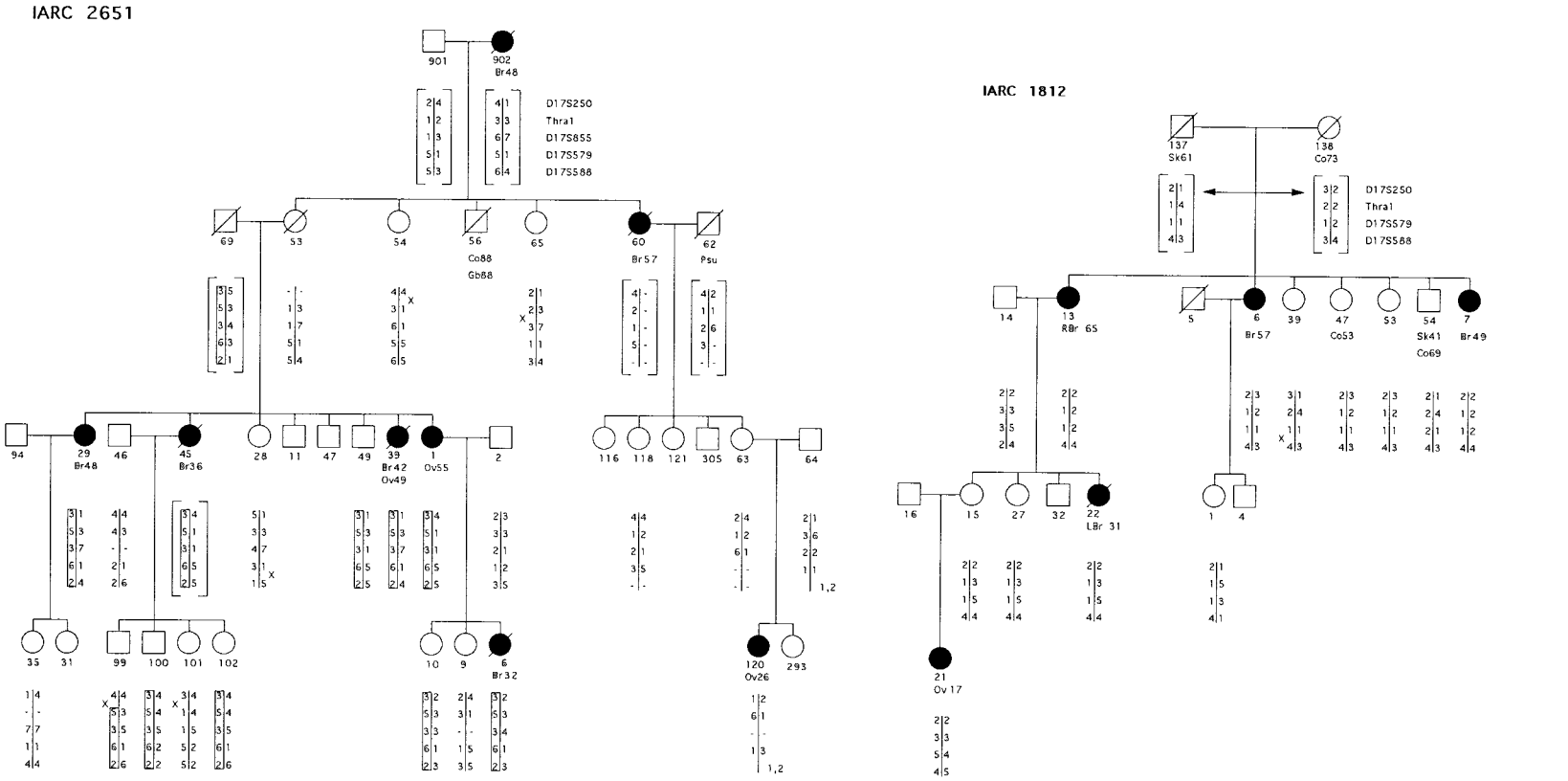


Figure 4 Pedigrees of IARC 2651 and IARC 1812. Symbols are as in figs. 1–3. Co = colon cancer; Gb = cancer of gall bladder; Psu = cancer, primary site unknown; and Sk = skin cancer.

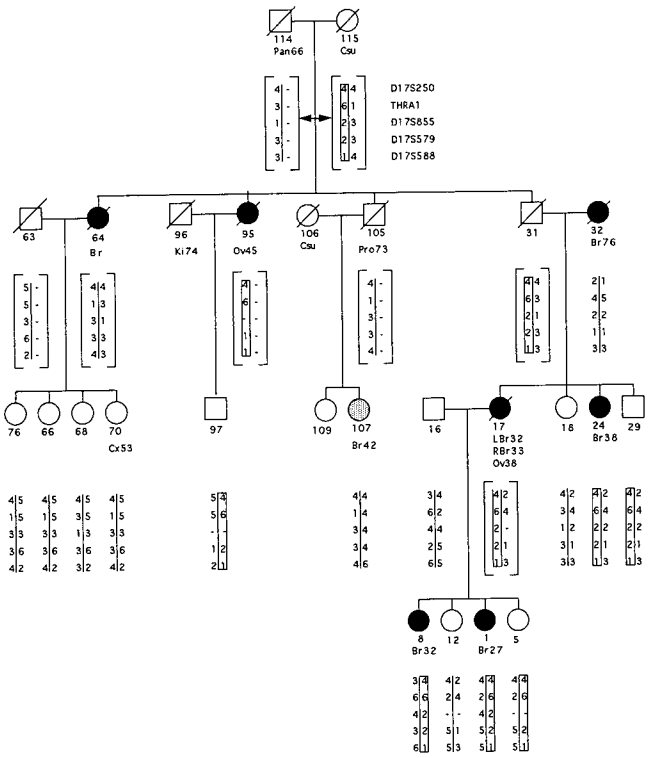


Figure 5 Pedigree IARC2090. Symbols are as in figs. 1–3. Pan = pancreatic cancer; Ki = kidney cancer; and Cx = cervical cancer. Individual 107 had in situ carcinoma.

tium families, and it will be important to extend these observations to other unlinked families.

In conclusion, we believe that, in addition to BRCA1, there are other genes that predispose to both cancer of the ovary and cancer of the breast. Although the overall evidence favors genetic heterogeneity, several explanations need to be explored before a particular family can be classified as unlinked. The extent of heterogeneity is much reduced for the subset of families with two or more cases of ovarian cancer. No doubt the situation will be clarified either when additional susceptibility loci are mapped or when BRCA1 is identified and the direct search for mutations in families becomes possible.

Acknowledgments

Groups and individuals contributing to this study are as follows: University of Aberdeen—N. Haites, B. Milner, and L. Allan; Canadian Breast Cancer Linkage Group—S. A. Narod, P. Tonin, J. Simard, J. Rommens, B. McGillivray, R. Green, E. Ives, N. Boyd, B. Rosen, D. Cole, K. Morgan, and R. Moslehi; Cancer Research Campaign, London and Cambridge—D. F. Easton, B. Ponder, J. Peto, S. Smith, D. Ford, M. Stratton, M. S. Piver, R. A. Dicioccio, H. Gallion, J. P. Struewing, and A. Whittemore; Christie Hospital and Holt Radium Institute, Manchester, United Kingdom—J. M. Birch, M. S. Santibanez-Koref, M. D. Teare, and D. G. Evans; French Cooperative Network—H. Sobol, D. Stoppa-Lyonnet, P. Lalle, Y. J. Bignon, C. Bonaiti, L. Essioux, C. Girodet, C. Maugard-Louboutin, G. Nicolleau, M. Longy, C. Toulouse, I. Holstein, D. Birnbaum, F. Eisinger, F. Karengueven, T. Noguchi, A. Hardouin, P. Rio, M. Machelard-Roumagnac, C. Nogues, O. Cohen-Haguenauer, A. Lortholary, and J.-O. Bay; University Hospital of Iceland—A. Arason, R. B. Barkdardottir and V. Egilsson; Imperial Cancer Research Fund, Leeds and London—D. T. Bishop, D. Kelsell, V. A. Murday, E. Solomon, N. Spurr, and G. Turner; International Agency for Research on Cancer—G. Lenoir, J. Feunteun, S. A. Narod, H. Lynch, J. Lynch, P. Watson, T. Conway, C. Bonnardel, O. Serova, and D. Torchard; Karolinska Hospital, Stockholm—A. Lindblom and C. Larsson; University of Leiden—P. Devilee, H. Vasen, I. van Leeuwen, R. S. Cornelis, and C. J. Cornelisse; Medical Research Council Human Genetics Unit, Edinburgh—M. Steel, D. Porter, E. Smyth, B. B. Cohen, and A. Carothers; University of Utah, Salt Lake City—L. A. Cannon-Albright, D. Goldgar, and M. Skolnick; German Cancer Research Center, Heidelberg—J. Chang-Claude, H. Becher, and U. Hamann; University of Lund—A. Borg and O. Johannsson; University of Michigan—B. Weber, F. Collins, and M. Boehnke; and Dana Farber Cancer Center—J. Garber and F. Li.

References

Amos CI, Shaw GL, Tucker MA, Hartge P (1992) Age at onset for familial epithelial ovarian cancer. JAMA 268:1896–1899

Anderson LA, Friedman L, Osborne-Lawrence S, Lynch E, Weissenbach J, Bowcock A, King M-C (1993) High density genetic map of the BRCA1 region of chromosome 17q12-q21. Genomics 17:618–623

Bewtra C, Watson P, Conway T, Read-Hippee C, Lynch HT (1992) Hereditary ovarian cancer: a clinicopathological study. Int J Gynecol Pathol 11:180–187

Biesecker BB, Boehnke M, Calzone K, Markel DS, Garber JE, Collins FS, Weber BL (1993) Genetic counselling for families with inherited susceptibility to breast and ovarian cancer. JAMA 269:1970–1974

Bowcock AM, Anderson LA, Friedman LS, Black DM, Osborne-Lawrence S, Rowell SE, Hall JM, et al (1993) THRA1 and D17S183 flank an interval of <4 cM for the breast-ovarian cancer gene (BRCA1) on chromosome 17q21. Am J Hum Genet 52:718–722

Buller RE, Anderson B, Connor JP, Robinson R (1993) Familial ovarian cancer. Gynecol Oncol 51:160–166

Chamberlain JS, Boehnke M, Frank TS, Kiousis S, Xu J, Guo S-W, Hauser ER, et al (1993) BRCA1 maps proximal to D17S579 on chromosome 17q21 by genetic analysis. Am J Hum Genet 52:792–798

Cornelis RS, Devilee P, Van Vliet M, Kuipers-Dijkshoorn N, Kersenmaeker A, Bardoel A, Khan PM, et al (1993) Allele loss patterns on chromosome 17q in 109 breast carcinomas indicate at least two distinct target regions. Oncogene 8:781–785

Easton DF, Bishop DT, Ford D, Crockford GP, Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Easton D, Ford D, Peto J (1993) Inherited susceptibility to breast cancer. Cancer Surv 18:95–113

Feunteun J, Narod SA, Lynch HT, Watson P, Conway T, Lynch J, Parboosingh J, et al (1993) A breast-ovarian cancer susceptibility gene maps to chromosome 17q21. Am J Hum Genet 52: 736–742

Foulkes W, Black D, Solomon E, Trowsdale J (1991) Allele loss on chromosome 17q in sporadic ovarian cancer. Lancet 338: 444–445

Goldgar DE, Cannon-Albright LA, Oliphant A, Ward JH, Linker G, Swensen J, Tran TD, et al (1993) Chromosome 17q linkage studies of 18 Utah breast cancer kindreds. Am J Hum Genet 52:743–748

Goldgar DE, Fields P, Lewis CM, Tran TD, Cannon-Albright LA, Ward JH, Swensen J, et al (1994) A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis. J Natl Cancer Inst 86:200–209

Greggi S, Genuardi M, Benedetti-Panici P, Cento R, Scambia G, Neri G, Mancuso S (1990) Analysis of 138 consecutive ovarian cancer patients: incidence and characteristics of familial cases. Gynecol Oncol 39:300–304

Hall JM, Friedman L, Guenther C, Lee MK, Weber JL, Black DM, King M-C (1992) Closing in on a breast cancer gene on chromosome 17q. Am J Hum Genet 50:1235–1242

Hall JM, Lee MK, Newman B, Morrow J, Anderson A, Huey B, King M-C (1990) Linkage analysis of early onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Jolly KW, Malkin D, Douglas EC, Brown TF, Sinclair AE, Look AT (1994) Splice-site mutation of the p53 gene in a family with hereditary breast-ovarian cancer. Oncogene 9:97–102

Kelsell DP, Black DM, Bishop DT, Spurr NK (1993) Genetic analysis of the BRCA1 region in a large breast/ovarian family: refinement of the minimal region containing BRCA1. Hum Mol Genet 2:1823–1828

King M-C, Rowell S, Love SM (1993) Inherited breast and ovarian cancer. what are the risks? what are the choices? JAMA 269:1975–1980

Lathrop GM, Lalouel JM (1984) Easy calculation of lod scores and genetic risks on small computers. Am J Hum Genet 36: 460–465

Lynch HT, Watson P, Conway TA, Lynch JF, Slominski-Carter SM, Narod SA, Feunteun J, et al (1993) Pilot study of DNA screening for breast/ovarian cancer susceptibility based on linked markers. Arch Intern Med 153:1979–1987

Narod SA (1993) Hereditary breast-ovarian cancer: how can we use the new DNA markers to improve patient management? Clin Invest Med 16:314–317

Narod SA, Feunteun M, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-23. Lancet 338:82–83

Narod SA, Madlensky L, Bradley L, Cole D, Tonin P, Rosen B, Risch HA (1994) Hereditary and familial ovarian cancer in southern Ontario. Cancer 74:2341–2346

Narod S, Tonin P, Lynch H, Watson P, Feunteun J, Lenoir G (1994) Histology of BRCA1-associated ovarian tumors. Lancet 343:236

Ott J (1985) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

Piver MS, Baker TR, Jishi MF, Sandecki AM, Tsukada Y, Natarajan N, Mettlin CJ, et al (1993) Familial ovarian cancer: a report of 658 families from the Gilda Radner Familial Ovarian Cancer Registry 1981–1991. Cancer 71:582–588

Saito H, Inazawa J, Saito S, Kasumi F, Koi S, Satoru S, Ryuichi K, et al (1993) Detailed deletion mapping of chromosome 17q in ovarian and breast cancer: 2-cM region on 17q21.3 often and commonly deleted in tumors. Cancer Res 53:3382–3385

Simard J, Feunteun J, Lenoir GM, Tonin P, Normand T, The VL, Vivier A, et al (1993) Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12-21: exclusion of candidate genes EDH17B2 and RARA. Hum Mol Genet 2:1193–1199

Smith CAB (1961) Homogeneity test for linkage data. Proc Soc Int Congress Hum Genet 1:212–213

Smith SA, Di Cioccio RA, Struewing JP, Easton DF, Gallion HH, Albertsen H, Mazoyer S, et al (1994) Localisation of the breast-ovarian cancer susceptibility gene (BRCA1) on 17q12-21 to an interval of < 1cM. Genes Chromosom Cancer 10:71–76

Smith SA, Easton DF, Ford D, Peto J, Anderson K, Averill D, Stratton M, et al (1993) Genetic heterogeneity and localizaion of a familial breast-ovarian cancer gene on chromosome 17q12-q21. Am J Hum Genet 52:767–776

Sobol H, Mazoyer S, Narod SA, Smith SA, Black DM, Kerbrat P, Jamot B, et al (1992) Genetic heterogeneity of early-onset familial breast cancer. Hum Genet 89:381–383

Steichen-Gersdorf E, Gallion HH, Ford D, Girodet C, Easton DF, DiCioccio RA, Evans G, et al (1994) Familial site-specific ovarian cancer is linked to *BRCA1* on 17q12-21. Am J Hum Genet 55:870–875

Stratton MR, Ford D, Neuhasen S, Seal, Wooster R, Friedman LS, King M-C, et al (1994) Familial male breast cancer is not linked to the BRCA1 locus on chromosome 17q. Nature Genet 7:103–107

Tonin P, Serova O, Simard J, Lenoir G, Feunteun J, Morgan K, Lynch H, et al (1994) The gene for hereditary breast-ovarian, BRCA1 maps distal to EDH17B2 in chromosome region 17q12-21. Hum Mol Genet 3:1679–1682

Varesco L, Caligo MA, Simi P, Black DM, Nardini V, Casarino L, Rocchi M, et al (1992) The NM23 gene maps to human chromosome band 17q22 and shows a restriction fragment polymorphism with Bgl II. Genes Chromosom Cancer 4:84–88

Weber JM (1990) Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. Nucleic Acids Res 18:4640

activity in human cells and tissues. Biochim Biophys Acta 1144:177–183

Van den Bogert C, Pennings A, Dekker HL, Luciakova K, Boezeman JBM, Sinjorgo KCM (1991) Quantification of mitochondrial proteins in cultured cells by immuno-flow cytometry. Biochim Biophys Acta 1097:87–94

Van Kuilenburg ABP, Dekker HL, Van den Bogert C, Nieboer P, Van Gelder BF, Muijsers AO (1991) Isoforms of cytochrome *c* oxidase: subunit composition and steady-state kinetic properties. Eur J Biochem 199:615–622

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AMS, Elsas LJ, et al (1988) Mitochondrial DNA mutations associated with Leber's hereditary optic neuropathy. Science 242:1427–1430

Address for correspondence and reprints: Dr. Roelof-Jan Oostra, The Netherlands Ophthalmic Research Institute, Department of Ophthalmogenetics, P.O. Box 12141, 1100 AC Amsterdam, The Netherlands.

0002-9297/95/5704-0026$02.00

Am. J. Hum. Genet. 57:957–958, 1995

Genetic Heterogeneity of Breast-Ovarian Cancer Revisited

To the Editor:

We have recently reported the results of a linkage analysis of 145 breast-ovarian cancer families (Narod et al. 1995). Each family has three or more cases of early-onset breast cancer (age <60) or of ovarian cancer, and all families have at least one case of ovarian cancer (there were nine site-specific ovarian cancer families). Overall, we estimated that 76% of the families were linked to the BRCA1 locus.

Among the 145 families studied, there were only 10 families that gave multipoint lod scores <-0.80, with markers flanking BRCA1, and that therefore were considered unlikely to be linked to this locus. Since the submission of our report, the BRCA1 gene has been identified (Miki et al. 1994), and the BRCA2 locus has been mapped to chromosome 13q (Wooster et al. 1994). Additional studies have now revealed the genetic basis of each of these 10 families—3 families were found to carry BRCA1 mutations, and 7 families show positive linkage to the BRCA2 locus.

IARC 2850 carries a BRCA1 mutation leading to the loss of exon 5. This family originally gave a multipoint lod score of -1.71 at BRCA1, but it can now be seen that the family contains two sporadic cases of cancer and that two recombination events have occurred between the flanking markers D17S250 and D17S579. The second family with a BRCA1 mutation is CRC 128. This mutation in this family leads to premature chain termination at codon 285. Because of the presence of a sporadic case of breast cancer in this family, at age 35 years, the lod score to BRCA1 is -0.91. Family RUL49 is a Dutch family with four early-onset breast cancers, one male breast cancer, and three ovarian cancers. Recently the family has been found to carry a BRCA1 frameshift mutation leading to termination at codon 780 (Hogervorst et al. 1995). Because of a sporadic early-onset breast cancer, the BRCA1 lod score is negative (-1.11). This is a rare example of a family with male breast cancer and a BRCA1 mutation. A second example has been reported by Struewing et al. (1995).

The other seven families appear to be linked to BRCA2, with multipoint lod scores ranging from 0.93 to 3.70 (table 1). Five of these families contain cases of male breast cancer, and five contain only a single case of ovarian cancer. For six families the evidence of linkage to BRCA2 is very strong (lod scores >1.6); for the other family (ICELAND 2204) the linkage evidence is more modest.

In summary, none of the 145 families in this large data set now provides clear evidence against both linkage to BRCA1 and linkage to BRCA2. These results indicate that BRCA1 and BRCA2 account for the majority of breast-ovarian cancer families. Studies are now underway that will estimate the proportions of breast cancer families attributable to each of these loci. It remains possible that a third breast cancer locus will be found, but it is unlikely that this locus will account for a significant proportion of families with breast and ovarian cancer.

STEVEN NAROD,[1] DEBBIE FORD,[2] PETER DEVILEE,[3] ROSA B. BARKARDOTTIR,[4] JORUNN EYFJORD,[5] GILBERT LENOIR,[6] OLGA SEROVA,[6] DOUGLAS EASTON,[2] DAVID GOLDGAR,[7] AND THE BREAST CANCER LINKAGE CONSORTIUM

[1]*Department of Medicine, Division of Medical Genetics, McGill University, Montreal;* [2]*Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey;* [3]*Department of Pathology and Human Genetics, University of Leiden, Leiden;* [4]*Department of Pathology, University Hospital of Iceland, and* [5]*Molecular and Cell Biology Research Laboratory, The Icelandic Cancer Society, Reykjavik;* [6]*International Agency for Research on Cancer, Lyon; and* [7]*Genetic Epidemiology Group, University of Utah, Salt Lake City*

References

Hogervorst FBL, Cornelis RS, Bout M, van Vliet MV, Oosterwijk JC, Olmer R, Bakker B, et al (1995) Rapid detection of BCRA1 mutations by the protein truncation test. Nat Genet 10:208–212

Table 1

Breast Ovary Families with Negative Lod Scores to BRCA1

Family	No. of Cancers: Breast, Female	Breast, Male	Ovarian	Lod Score to BRCA1[a]	BRCA1 Mutation?[b]	Lod Score to BRCA2[b,c]
IARC 2850	15	0	2	−1.71	Yes	...
RUL 49	4	1	3	−1.11	Yes	...
CRC 128	4	0	4	−.91	Yes	...
CRC 186	16	1	1	−2.61	...	3.70
RUL9	7	0	1	−2.01	...	1.64
UTAH 107	32	3	6	−2.45	...	3.48
UTAH 2044	7	1	4	−1.32	...	2.11
ICELAND 6	15	1	1	−.96	...	2.92
ICELAND 2204	3	4	1	−.85	...	.93
ICELAND 80004	8	0	1	−1.89	...	1.82

NOTE.—The linkage model is presented in the report by Narod et al. (1995).
[a] Based on the markers D17S250 and D17S579.
[b] An ellipsis indicates that the test was not done.
[c] Based on markers D13S260, D13SS289, and D13S267.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal FA, Harshman K, Tavrigian S, Liu Q, et al (1994) A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66–71

Narod SA, Ford D, Devilee, P, Barkardottir RB, Lynch HT, Smith SA, Ponder BAJ, et al (1995) An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Am J Hum Genet 56:254–264

Struewing JP, Brody LC, Erdos MR, Kase RG, Giambarresi TR, Smith SA, Collins FS, et al (1995) Detection of eight *BRCA1* mutations in 10 breast/ovarian cancer families, including 1 family with male breast cancer. Am J Hum Genet 57:1–7

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, et al (1994) Localisation of a breast cancer susceptibility gene BRCA2, to chromosome 13q12-13. Science 265:2088–2090

Address for correspondence and reprints: Dr. Steven Narod, Montreal General Hospital, Division of Medical Genetics, 1650 Cedar Avenue, Room L10-107, Montreal, Quebec H3G 1A4, Canada.

0002-9297/95/5704-0027$2.00

Am. J. Hum. Genet. 57:958–960, 1995

CFTR Gene Variant for Patients with Congenital Absence of Vas Deferens

To the Editor:

Obstructive azoospermia due to congenital absence of vas deferens is a prominent clinical feature among male patients with cystic fibrosis (CF) (Holsclaw et al. 1971). A similar autosomal recessive condition with no other CF manifestations is classified as congenital bilateral absence of vas deferens (CBAVD; McKusick 277180; Schellen and van Stratten 1980). Since 50%–64% of CBAVD patients have been found to be positive for at least one known CFTR mutation, it is believed that at least part of the CBAVD population represents an atypical form of CF affecting only the male reproductive system (Dumur et al. 1990; Anguiano et al. 1992; Gervais et al 1993; Osborne et al. 1993; Patrizio et al. 1993; Culard et al. 1994). This explanation is not completely satisfactory, however, because only ~10% of CBAVD patients are found to carry known CF mutations on both chromosomes, even after exhaustive screening of the entire CFTR coding region. Here we present data to show that a previously known sequence variant in intron 8 of the CFTR gene (Chu et al. 1993) is a specific and frequent mutation associated with CBAVD.

Varied lengths of a thymidine (T)–tract (5, 7, or 9T) have been noted in front of the splice-acceptor site of intron 8 (Chu et al. 1993). The length appears to correlate with the efficiency of exon 9 splicing, with the 5T variant that is present in 5% of the CFTR alleles among the Caucasian population producing almost exclusively (95%) exon 9-minus mRNA (Chu et al. 1993). The effect of this T-tract polymorphism in CFTR gene expression is also documented by its relationship with a known CF mutation R117H (Dean et al. 1990): While R117H(5T) is found in typical CF patients with pancreatic sufficiency, R117H(7T) is associated with CBAVD (Kiesewetter et al. 1993). More recently, CFTR alleles

INTERNATIONAL JOURNAL OF CANCER (IN PRESS)

RISK MODIFIERS IN CARRIERS OF BRCA1 MUTATIONS

Steven A NAROD[1], David GOLDGAR[2], Lisa CANNON-ALBRIGHT[2], Barbara WEBER[3], Roxanna MOSLEHI[4], Elizabeth IVES[5], Gilbert LENOIR[6], Henry LYNCH[7]

[1] Departments of Medicine and Human Genetics, McGill University, Montreal, Canada, 1650 Cedar Avenue, Montreal, Canada, H3G 1A4; [2] Genetic Epidemiology Group, University of Utah, Salt Lake City, Utah; [3] Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania; [4] Department of Medical Genetics, University of British Columbia, Vancouver, BC; [5] Faculty of Medicine, Memorial University of Newfoundland, St. John's; [6] International Agency for Research on Cancer, Lyon France; [7] Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, Nebraska.

(Running title) Risk Modifiers in BRCA1 Carriers
Narod et al.

Correspondence to: Dr Steven Narod, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec, Canada, H3G 1A4

Tel: (514) 937-6011 (local 4067) FAX: (514) 934-8273

SUMMARY The majority of, but not all, women with mutations in the BRCA1 gene will be affected with breast or ovarian cancer by the age of 70. To establish whether known risk factors modify susceptibility to cancer in these women we have studied the reproductive histories of 333 North American women who were found by haplotype analysis to carry BRCA1 mutations. An increased risk for breast cancer was associated with low parity and with recent birth cohort. The risk of ovarian cancer decreased with increasing age at last childbirth; however, in contrast to the case for sporadic cancer, the risk of ovarian cancer in BRCA1 carriers was found to increase significantly with increasing parity.

INTRODUCTION

The BRCA1 gene on chromosome region 17q12 predisposes to both breast and ovarian cancer (Hall et al, 1990, Narod et al, 1991, Miki et al, 1994. It has been estimated that between 1 in 200 and 1 in 2000 women carry mutations in this gene (Easton et al, 1994). Carriers of BRCA1 mutations in breast-ovary cancer families can be identified by performing linkage analysis using polymorphic markers from chromosome 17q (e.g. Easton et al, 1993; Narod et al, 1995) or directly by genetic sequencing of constitutional DNA (Miki et al, 1994; Simard et al, 1994; Friedman et al, 1994; Castilla et al, 1994; Shattuck-Eidens et al, 1995).

The clinical expression of BRCA1 appears to vary from family to family (Easton et al, 1994). These differences may prove to be due in part to different mutant BRCA1 alleles associated with different degrees of penetrance. The majority of known carriers to date have been identified in large kindreds with multiple affected members. It has been suggested that these extended pedigrees may be atypical, and that BRCA1 mutations with less profound effects may be more common (Rowell et al, 1994). Within a family it is also common to see different effects of a BRCA1 mutation; some women may be affected with breast cancer in their thirties, whereas others may remain unaffected into their seventies. There are several alternate explanations for the observed variability within families, including: 1) the random accumulation of mutations in other critical genes; 2) possible unidentified environmental factors; 3) the modification of cancer risk by pregnancy or by the use of exogenous hormones; and 4) the influence of allelic variation of other genes on the BRCA1 phenotype.

The discovery of BRCA1 modifiers may clarify our understanding of the mode of action of this gene and may lead us to refine our current models of hormonal carcinogenesis. It is also important to identify non-genetic risk modifiers in order to

counsel BRCA1 mutation carriers about potential means of reducing their risk. We have constructed a historical cohort of 333 female carriers of BRCA1 mutations in four centers in North America and have studied the effects of reproductive and other variables on breast and ovarian cancer incidence.

SUBJECTS AND METHODS

Identification of families Families were identified through four clinical centers in North America where BRCA1 mutations analysis is currently offered. All families contained cases of ovarian cancer in addition to multiple cases of breast cancer. No family contained a case of male breast cancer. It is estimated that 88% of such families are linked to BRCA1 (Narod et al, 1995). All families had lod scores of 0.5 or greater with linkage to BRCA1, using either flanking markers (D17S250, THRA1, D17S579, D17S588) or the intragenic marker D17S855. Details of the markers used are provided in Miki et al (1994) and Easton et al (1993). A lod score of 0.50 corresponds to a posterior probability of linkage of 95%, but for most families the probability of linkage to BRCA1 was much higher than this. Mutations in BRCA1 have now been identified for 21 of the 28 families, including all 3 of 3 Utah families, 4 of 6 McGill families 2 of 3 Pennsylvania families, (Simard et al, 1994; Friedman et al, 1994; Shattuck-Eidens et al, 1995) and 12 of 16 Creighton families (unpublished data).

Identification of BRCA1 carriers Individuals were considered to be carriers in the families by the inspection of segregating haplotypes. Women for whom haplotype data was not available were excluded from the analysis. In some cases, haplotypes were constructed by inference using marker data derived from children and spouses. Women were considered to be carriers of BRCA1 mutations if they inherited the allele of risk for the intragenic marker D17S855. If D17S855 was not typed, or was uninformative, the flanking markers were used to establish the carrier status. If there was evidence for a recombination event between the flanking markers THRA1 and D17S579, and the position of the recombinant could not be resolved with respect to D17S855, the BRCA1 status was considered to be ambiguous, and the case was excluded from the analysis. Within sibships BRCA1 testing was was independent of disease status.

A total of 333 BRCA1 carrier women were identified through haplotype analysis of 28 breast-ovary cancer families using chromosome 17 markers. Of these, 169 women had been diagnosed with breast cancer by 1994 (mean age of diagnosis 40.9 years), 72 had ovarian cancer (53.4 years) and 25 women had cancers of both types (table I). 98 women are currenty alive and unaffected (mean age 38.4 years) and 19 have died of another cause. The years of birth of the women ranged from 1859 to 1973. Women were followed, on average, until the age of 50.0 years. The

ratio of cases of breast cancer to ovarian cancer was 2.3 to 1 overall; this ratio varied from 1.90 (McGill) to 3.83 (Pennsylvania).

Information on reproductive history was routinely obtained on living study subjects at the time blood was drawn for the genetic analysis, either by written questionnaire or by telephone interview. This information included their year of birth and the number and ages of all pregnancies. Information for the 117 deceased women was obtained from a close relative, usually a sister or a daughter. Age of menarche was only available for the McGill and Creighton families.

Statistical analysis The data was analysed by survival analysis methods. Individuals were considered to be at risk of breast or ovarian cancer from the age of twenty until: 1) the development of breast or ovarian cancer; 2) death from another cause; 3) prophylactic surgical removal of the breasts or ovaries; September 1994. The survival curves were constructed using Kaplan-Meir estimates. Differences in Kaplan-Meir curves were tested for statistical significance using the log-rank test. The effects of covariates on survival was modelled using Cox regression assuming a proportional hazards model (Cox, 1972) Nulliparous women were excluded from the univariate analyses of estimates of the effect of age of first birth and age of last birth. When multivariate analysis included variables for both age of first (or last) birth and parity, the age of first (or last) birth was coded as zero for nulliparous women.

RESULTS

The women in this study represent a historical cohort of 339 identified BRCA1 mutation carriers from 28 families, each with a minimum of one case of ovarian cancer and three cases of breast cancer. The age-specific incidence rates of breast and ovarian cancer in this cohort are presented in figure 1. The maximal breast cancer incidence was observed in the age group 41 - 45 years. Breast cancer rates declined modestly thereafter. The ovarian cancer incidence was relatively constant above the age of 45, and was greatest between ages 51 and 55. For women over the age of 50 the annual risk of ovarian cancer exceeded the risk for breast cancer. The cumulative incidence of breast cancer was 52.2% to age 50 (95% confidence interval; 45.9% - 58.4%) and 74.6% to age 70 (67.0% - 82.2%); for ovarian cancer these figures were 19.5% to age 50 (12.5% - 24.5%) and 56.3% to age 70 (45.9% - 66.7%).

When each of the predictive variables was considered on its own, increased risks of breast cancer were associated with menarche below the age of 12 years (relative risk = 1.57; $p = 0.053$), with parity of less than three (relative risk = 2.04; $p <$

0.001) and with year of birth after 1930 (relative risk = 2.72; p < 0.001). The cohort effect was surprisingly strong. It is estimated that 42.4% of women in the recent cohort will be diagnosed with breast cancer by age 40, as compared to 12% for women born before 1930.

Because of the possibility that the observed cohort effect may be biased by the means of ascertainment of the pedigrees, the analysis was repeated, but restricted to the largest single pedigree. Utah kindred 2082 was ascertained by linkage of the Utah Genealogy to the Utah cancer registry, based on a cluster of two breast and four ovarian cancers, and was independent of age-of onset of the breast cancers. The relative risk in this single pedigree associated with recent cohort was 3.84 (p = 0.013). Because of the large size of this pedigree (25 cases of breast cancer and 21 cases of ovarian cancer) the penetrance figures are unlikely to be influenced unduly by the effect of inclusion or exclusion of a small number of individuals.

The penetrance of breast cancer was not different for BRCA1 carriers who inherited the mutation from their father as for those who inherited it from their mother (p = 0.96).

Because of the effect of year of birth on penetrance, and because the risk for breast cancer varied between centers, the effects of the reproductive variables were studied while adjusting for study center and for birth cohort. Univariate and multivariate relative risks are presented in table II. Age of first birth, age of last birth and parity are inter-related, and a series of alternate models was evaluated in order to eliminate superfluous variables. In no model were age of first birth or age of last birth found to be independently predictive of breast cancer risk, and both these variables could be eliminated. In contrast, increasing parity was found to be protective for breast cancer (table III). A relative risk of 0.85 was estimated for each additional birth up to five or more (p = 0.002). The cumulative incidence of breast cancer in women with two or fewer births is compared to the incidence in women with three or more births in figure 2. There were notable trends in fertility patterns within this cohort. Women born after 1930 had, on average, fewer births (2.37 vs. 4.95) and an earlier average age of last birth (28.5 vs 34.5 years) than women born before 1930. However, the magnitude of the cohort effect was not diminished by adjustment for the reproductive variables (table II).

The best-fitting model for predicting ovarian cancer risk also included parity, but the direction of the risk associated with multiple births was opposite to that found for breast cancer. Each birth conferred an additional 40% increase in risk (relative risk = 1.40; p = 0.005) for up to five births. Although women with multiple parity were at increased ovarian cancer risk, there was also a protective effect of a late birth. When

analysed as a continuous variable in the multivariate model, each five-year interval in age of last birth was associated with a risk reduction of 18% (relative risk = 0.82; p = 0.03). Compared with women who last gave birth at age 29 or earlier, those with a last birth at age 30 or above were at 48% less risk. Women who had all their children over the age of thirty, or who were nulliparous formed a low risk group for ovarian cancer (relative risk = 0.30, p = 0.03).

As was the case for breast cancer, the risk for ovarian cancer was much higher in the recent birth cohort (relative risk = 2.08 for women born after 1930; p = 0.007). To eliminate the possible confounding effect of oral contraceptive use in recently born women (oral contraceptives are associated with low parity and with reduced ovarian cancer risk), to minimize the cohort effect, and to dissociate the effect of ovarian cancer on subsequent parity, the survival analysis was repeated, but was restricted to the subgroup of women who were born before 1930, and who were cancer-free at age 35 (figure 3). The results of the multivariate model in this restricted analysis were essentially unchanged - high parity continued to be predictive of ovarian cancer risk (relative risk = 1.95 for each additional birth up to five; p = 0.00028) and late birth was also protective in this subgroup (relative risk = 0.68 per five year interval; p = 0.005).

DISCUSSION

Several observations in this historical cohort suggest the interaction of environmental and genetic factors, including the incomplete penetrance of the BRCA1 gene and the trend towards earlier diagnosis. Reproductive factors do not appear to be sufficient to explain the observed variability in gene expression and penetrance.

In our study we estimated a risk of 75% to age 70 for breast cancer and 56.3% for ovarian cancer. It is perhaps surprising that the age-specific risks for ovarian cancer exceed the risks for breast cancer in BRCA1 carriers above the age of 50. Because each family was selected to contain at least one case of ovarian cancer, these penetrance estimates are pertinent to families with the breast-ovarian cancer syndrome. The total number of breast cancers outnumber ovarian cancers by 2.3 to 1 overall, but the majority (81.6%) of cases of breast cancer occur before the age of 50, and many of these women did not survive the initial cancer. Also, we did not consider women to be at risk for ovarian cancer after oophorectomy. Oophorectomy is used as a prophylactic measure and may also be used in the treatment of breast cancer. Our risks are similar to those of the Breast Cancer Linkage Consortium (BCLC) who estimate risks of 87% for breast cancer, and 44% for ovarian cancer to age 70 (Easton et al, 1994; Easton et al, 1995). Our estimate of

ovarian cancer penetrance is higher, but each of our families was selected because it contains at least one case of ovarian cancer. BRCA1 mutations have also been identified in families with site-specific breast cancer, but in our experience this group represents a minority of the families with mutations.

Goldgar et al (1994) found that the sex of the carrier parent in family 2082 was significantly associated with the type of cancer observed in the child (breast versus ovarian). We found no influence of the sex of the transmitting parent on the penetrance of either type of cancer in our extended series of pedigrees.

In this cohort of BRCA1 carriers the risk of cancer of either type was much higher in women born after 1930. The effect was equally strong using cut-off years of 1920 (i.e. relative risk of 2.68 for breast cancer) or of 1940 (relative risk 3.06 for breast cancer). We do not believe that the trend towards earlier diagnosis can be explained by better screening in the recently born women. It is our experience that most tumors have been identified by the detection of a palpable mass by the patient herself or by her physican. There has been no improvement in life expectancy consistent with extensive screening; in the earlier cohort survival was 80.0% to age 50, and in the recent cohort suvival to age 50 is 74.9%. Mammography data was not routinely available for the women in all families, and this will be the subject of a future prospective study.

Bias in the ascertainment of pedigrees could lead to spurious cohort effects, in particular if recently diagnosed early-onset cases were to lead to the selective inclusion of families in the study. We do not feel that our sampling strategy has produced a biased result for several reasons. The penetrance of breast cancer was significantly greater in the recent cohort in all four centers. The cohort effect was clearly present in the single large Utah pedigree, Kindred 2082 (see results). The Creighton families have been ascertained over the past 25 years and many of the recent breast cancer cases were diagnosed following the original ascertainment. Ten of the 20 cases of cancer in the largest family (kindred 1816) were diagnosed since 1978, at which time the pedigree was ascertained. A spurious temporal effect could also result if women were preferentially included in the recent cohort because they were affected. All women at risk in these 28 families were asked to participate in the genetic study, whether or not they were affected with cancer, and in most cases, the sampling of women in the relevant sibships was complete. Only women for whom DNA testing revealed the BRCA1 carrier haplotype, directly or by inference, were included. The proportion of subjects in the recent cohort who were affected (57.1%) was in fact lower than the proportion of affected women in the earlier cohort (79.2%). Preferential sampling of affected women should lead to an overestimate of penetrance. Our estimates of breast cancer penetrance is slightly lower than the

BCLC figure, which was derived using other methods (Eastone et al, 1995). The reason for the observed cohort effect is unclear, but does not seem to be explainable by the trends in fertility. Similar patterns of increasing risk in families with dominant cancer syndromes have been reported for familial malignant melanoma (Battistutta et al, 1994) and the Li-Fraumeni syndrome (L. Strong, personal communication).

The risk of breast cancer declined with increasing parity in BRCA1 carriers, in keeping with studies of breast cancer in the general population (Brinton et al, 1982; Rosner et al, 1994) but an early age of first birth did not confer additional protection. Giving birth at a young age has been observed to protect against breast cancer in the general population in several studies (Brinton et al, 1982; Rosner et al, 1994).

Alternate models proposed to explain the epidemiology of sporadic ovarian cancer include an increased risk from incessant ovulation or from prolonged exposure to raised levels of gonadotropins (Whittemore et al, 1992). Neither of these models is consistent with the pattern we observe in hereditary ovarian cancer. Pregnancy is associated with a reduction in the number of ovulations and with decreased gonadotropin levels, but is associated with an increased risk of ovarian cancer in BRCA1 carriers. The risk of sporadic ovarian cancer is reduced after giving birth (Adami et al, 1994). The protection offered by a late birth in BRCA1 carriers is consistent with the hypothesis of pregnancy-dependent clearance of malignant ovarian cells raised by Adami et al (1994). Multiparity is protective against sporadic ovarian cancer (Adami et al, 1994) but appears to be a risk factor in BRCA1 carriers. In a previous study of the Utah population database (Kerber and Slattery, 1995) it was found that giving birth to six or more children did not protect against familial ovarian cancer (odds ratio = 1.11), but was strongly protective against non-familial ovarian cancer (odds ratio = 0.29).

Oral contraceptive use reduces the risk of ovarian cancer in the general population (Parazzinni et al, 1991) and deserves consideration as a potential preventive therapy, but it is not yet clear if the effect of oral contraceptives on BRCA1-associated ovarian cancer will be in the same direction. It is of equal importance to establish if the use of oral contraceptives influences the incidence of breast cancer in BRCA1 carriers, especially in light of our unexplained observation of increased rates of cancer among recent generations of BRCA1 carriers.

Genetic counselling and risk assessment by using linked markers and by direct mutation detection is now offered in our centers and elsewhere. It is common for women to ask about the impact of the known risk factors on hereditary breast and ovarian cancer, and how they may modify their personal risk. The

absolute risk of breast cancer in these women declined after the time of the menopause, suggesting that ovarian hormones may be important in promoting tumour development in BRCA1 carriers. The effect of prophylactic oophorectomy or of tamoxifen use in preventing BRCA1-associated breast cancer, and the safety of hormone replacement therapy deserve particular attention in controlled follow-up studies. Our results indicate that it is inadequate to extrapolate from observations of sporadic cancers and that studies of the hereditary subgroups are required.

Table I. Families used in the Analysis

Centre	Families	Total Carriers	Number with Breast Cancer	Number with Ovarian Cancer	Unaffected Carriers
Creighton	16	162	87	36	53
Utah	3	103	40	19	47
McGill	6	34	19	10	8
Pennsylvania	3	34	23	7	9
TOTAL	28	333	169	72	117

Table I, Continued

	Parity	Average age of first birth	Average age of last birth
Creighton	2.93	22.9	30.0
Utah	3.94	23.1	32.1
McGill	2.67	23.1	31.6
Pennslyvania	3.03	22.1	28.8
TOTAL	3.22	22.9	30.7

Table II. Cox proportional hazards modelling of the effect of reproductive variables on breast cancer incidence; complete model.

Variable	Cancer	Univariate Relative Risk	p	Multivariate Relative Risk	p
COHORT	Breast	2.75	<0.001	2.53	<0.001
	Ovary	2.37	0.003	2.17	0.01
AGE OF FIRST BIRTH	Breast	0.94	0.43	0.98	0.85
	Ovary	0.64	0.007	1.18	0.32
AGE OF LAST BIRTH	Breast	0.76	<0.001	1.02	0.85
	Ovary	0.70	0.003	0.73	0.03
PARITY	Breast	0.78	<0.001	0.84	0.06
	Ovary	1.05	0.54	1.47	0.003

Definitions of variables: COHORT: born after 1930 versus born before or in 1930; AGE OF FIRST BIRTH: Relative risks for each five-year increment in age of first birth; AGE OF LAST BIRTH: Relative risks for each five-year increment in age of last birth; PARITY: Relative risks are for each additional birth up to five, compared to nulliparous women. Relative risks are adjusted for other variables in the table as well as for study centre.

Table III. Cox proportional hazards modelling of the effect of reproductive variables on breast cancer incidence: final model.

Variable	Cancer	Multivariate Relative Risk	p-value
COHORT	Breast	2.42	<0.001
	Ovary	2.07	0.007
AGE OF LAST	Breast	****	****
BIRTH	Ovary	0.82	0.03
PARITY	Breast	0.85	0.002
	Ovary	1.40	0.005

Definitions of variables as in table II. Relative risks are adjusted for other variables in the table and for study center. ***** Age of last birth was not included in the model for breast cancer.

LEGENDS FOR FIGURES

Figure 1. Incidence of breast and ovarian cancer in cohort of BRCA1 carriers.

Figure 2. Influence of parity on breast cancer incidence.

Figure 3. Influence of parity on ovarian cancer incidence for women born before 1930. All women cancer-free at age 35.

REFERENCES

ADAMI, H.O., HSIEH, C.C., LAMBE, M., TRICHOPOLOUS, D., LEON, D., PERSSON, I., EKBOM, A. and JANSON. P.O. Parity, age at first childbirth, and risk of ovarian cancer. Lancet, 40, 1250-54 (1944).

BATTISTUTTA, D., PALMER, J., WALTERS, M., WALKER, G., NANCARROW, D. and HAYWARD, N. Incidence of familial melanoma and MLM2 gene. Lancet 344, 1607-1608 (1994).

BRINTON, L., HOOVER, R. and FRAUMENI, J.F. Interaction of familial and hormonal risk factors for breast cancer. J. Natl. Cancer Inst. 69, 817-822 (1982).

CASTILLA, L.H., COUCH, F.J., ERDOS, M.R., HOSKINS K.F., CALZONE, K., GARBER, J.E., BOYD J., LUBIN, M.B., SESHANO, M.L., BRODY, L.C., COLLINS F.S. and WEBER, B.L. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. Nature Genetics 8, 387-391 (1994).

COX, D.R. Regression models and life tables. J R Stat Soc 1972; 34, 187-220 (1972).

EASTON, D.F., NAROD, S.A., FORD, D. and STEEL, M. The genetic epidemiology of BRCA1. Lancet 344, 761 (1994).

EASTON, D.F., BISHOP, D.T., FORD, D., CROCKFORD, G.P. and the Breast Cancer Linkage Consortium. Genetic linkage analysis in familial breast and ovarian cancer - results from 214 families. Amer. J. Hum. Genet. 52, 678-701 (1993).

EASTON, D.F., FORD, D., BISHOP, D.T. and the Breast Cancer Linkage Consortium. Breast and ovarian cancer incidence in BRCA1 mutations carriers. Am. J. Hum. Genet. 56, 265-271 (1995).

FRIEDMAN, L.S., OSTERMEYER, E.A., SZABO, C.I., DOWD, P., LYNCH,

E.D., ROWELL, S.E. and KING. M.-C.. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nature Genetics 8, 399-404 (1994).

GOLDGAR, D.E., FIELDS, P., LEWIS, C.M., TRAN T. D., CANNON-ALBRIGHT, L., WARD, J.H., SWENSEN, J., SKOLNICK, M.H. A large kindred with 17-q linked breast and ovarian cancer: genetic, phenotypic and genealogical analysis. J. Natl. Cancer. Inst. 86, 200-209 (1994).

HALL, J.M., LEE, M.K., NEWMAN, B., MORROW, J., ANDERSON L., HUEY, B. and KING, M.-C. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250, 1684-1689 (1990).

KERBER, R.A. and SLATTERY, M.L. The impact of family history on ovarian cancer risk: the Utah population database. Arch. Int. Med. 155, 905-12 (1995).

MIKI, Y. and 44 others., A strong candidate gene for the breast and ovarian cancer susceptibility gene, BRCA1. Science 266, 66-71 (1944).

NAROD, S.A., FEUNTEUN, J., LYNCH, H.T., WATSON, P., CONWAY, T., LYNCH, J. and LENOIR.G.M. Familial breast-ovarian cancer locus on chromosome 17q12-23. Lancet 338, 82-83 (1991).

NAROD, S.A., and 27 others. An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Amer. J. Hum. Genet. 56, 254-264 (1995).

PARAZZINNI, F., FRANCHESI, S., LA VECCHIA, C. and FASOLI, M. The epidemiology of ovarian cancer. Gynecol Oncol 43: 9-23 (1991).

SIMARD, J. and 15 others, Common origins of BRCA1 mutations in Canadian breast and ovary cancer families. Nature Genetics 8:

392-398 (1994).

ROWELL, S., NEWMAN, B., BOYD, J. and KING, M.C. Inherited predisposition to breast and ovarian cancer. Am. J. Hum. Genet. 55: 861-865 (1994).

ROSNER, B., COLDITZ, G.A., WILLETT, W.C. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am. J. Epidemiol. 139, 819-35 (1994).

SHATTUCK-EIDENS, D., and 32 others. A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene: Implications for presymptomatic testing and screening. J. Amer. Medic. Soc. 273, 535-541 (1995).

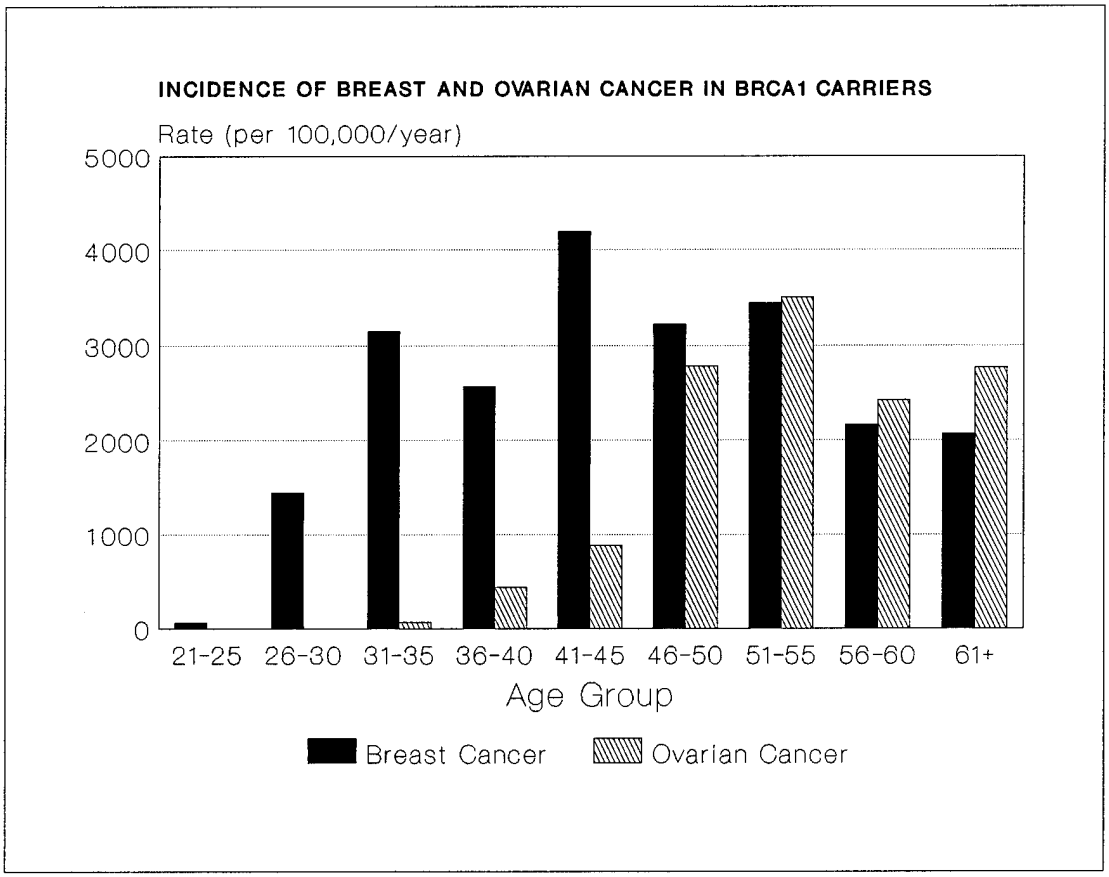
WEBER, B. Susceptibility genes for breast cancer. New. Engl. J. Med. 331, 1523-1524 (1994).

WHITTEMORE, A.S., HARRIS, R., INTYRE J. and the Collaborative Ovarian Cancer Group. Characteristics relating to ovarian cancer risk; collaborative analysis of 12 US case-control studies: IV The pathogenesis of epithelial ovarian cancer. Am. J. Epidemiol. 136, 1212-20 (1992).

WILLETT, W.C., STAMPFER, M.J., COLDITZ, G.A., ROSNER, B.A., HENNEKENS, C,H. and SPEIZER, F.E. Moderate alcohol consumption and the risk of breast cancer. New Engl. J. Med. 316, 1174 - 80 (1987).

ACKNOWLEDGEMENTS

This work was supported by the National Cancer Institute of Canada Breast Cancer Initiative, the National Instutes of Health grants 1R01 CA 63678-01, CA-57601, CA-61231, the Department of the Army grants DAMD17-94-J-4340 and DAMD17-94--J 4299, the State of Nebraska grant LB595 and the Canadian Genetics Diseases Network. SN is supported by the Fonds de la Recherche en Sante du Quebec. We thank Michael Pollak, William Foulkes, Kenneth Morgan and Harvey Risch for helpful discussions and Jean-Sebastien Brunet for exellent statistical support. We thank the families for their ongoing co-operation and for the clinicians and scientists who assisted in data collection and analysis, including P Tonin, J Simard, J Rommens, B McGillivray, R Green, N Boyd, B Rosen, D Cole (Canada), J Feunteun, J Lynch, P Watson, T Conway, C Bonnardel, O Serova (Creighton), D Anderson (Utah) and K Calzone (Pennsylvania).


INCIDENCE OF BREAST AND OVARIAN CANCER IN BRCA1 CARRIERS
Rate (per 100,000/year)
5000
4000
3000
2000
1000
0
21-25
26-30
31-35
36-40
41-45
46-50
51-55
56-60
61+
Age Group
Breast Cancer
Ovarian Cancer

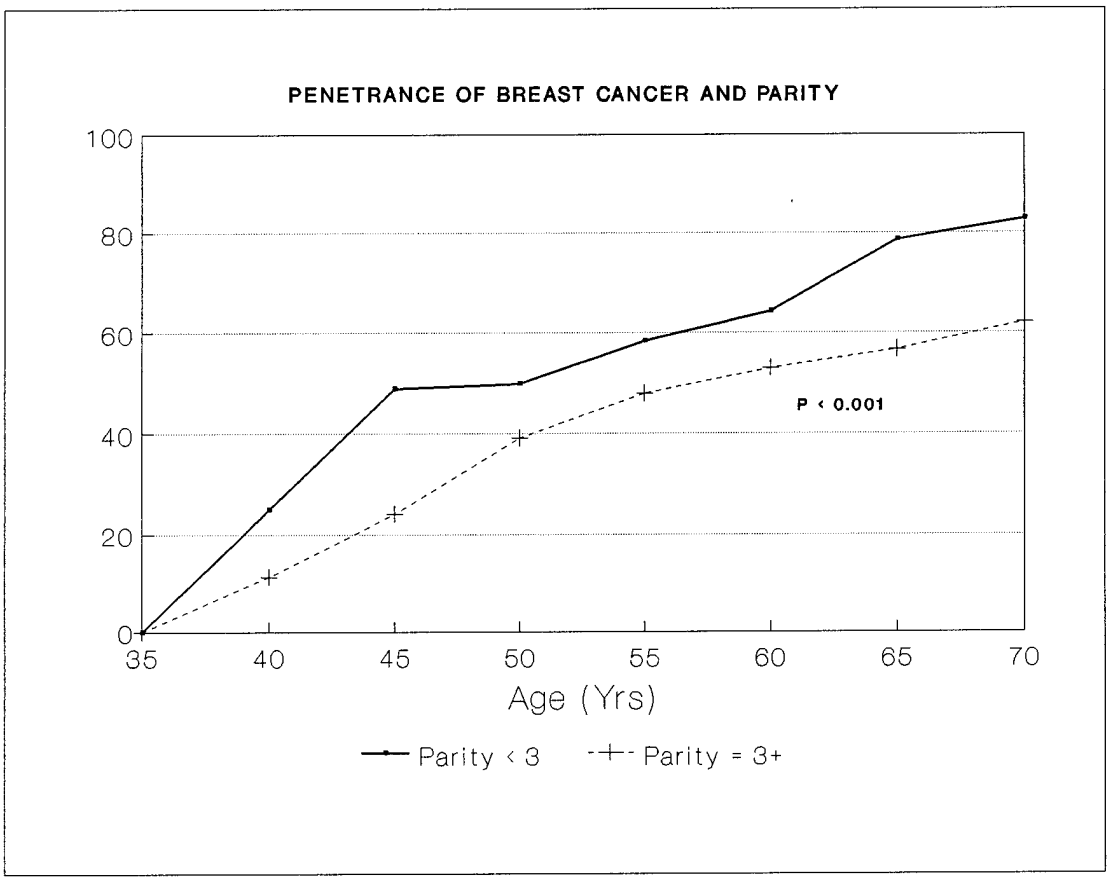
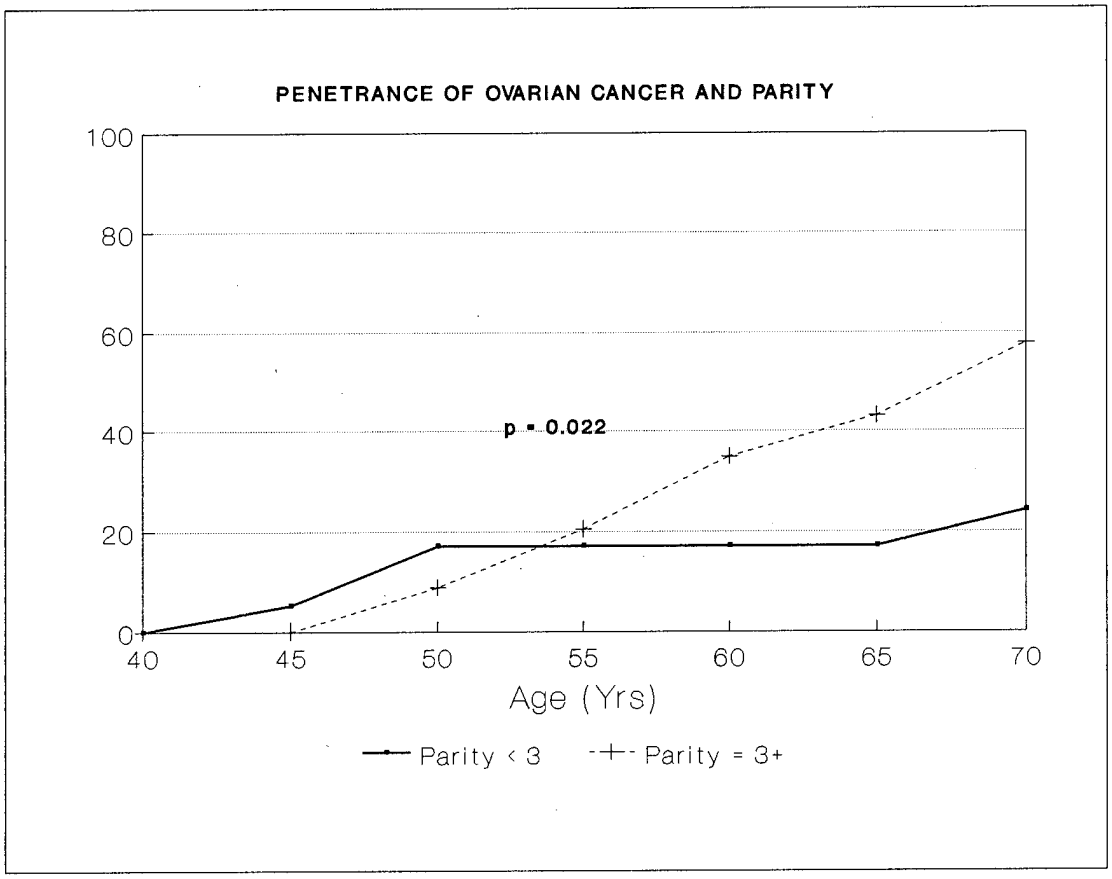

PENETRANCE OF BREAST CANCER AND PARITY
100
80
60
40
20
0
35
40
45
50
55
60
65
70
Age (Yrs)
P < 0.001
Parity < 3
Parity = 3+

FIG 6


PENETRANCE OF OVARIAN CANCER AND PARITY
100
80
60
40
20
0
p = 0.022
40
45
50
55
60
65
70
Age (Yrs)
Parity < 3
Parity = 3+

NAROD FIG 2

Human Molecular Genetics, 1994, Vol. 3, No. 11 1919–1926

A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene *BRCA1*

S.L.Neuhausen[1,*], J.Swensen[1], Y.Miki[1], Q.Liu[4], S.Tavtigian[4], D.Shattuck-Eidens[4], A.Kamb[4], M.R.Hobbs[2], J.Gingrich[5], H.Shizuya[6], U.-J.Kim[6], C.Cochran[7], P.A.Futreal[7], R.W.Wiseman[7], H.T.Lynch[8], P.Tonin[9], S.Narod[9], L.Cannon-Albright[3], M.H.Skolnick[1,4] and D.E.Goldgar[1]

[1]Department of Medical Informatics, [2]Department of Human Genetics and [3]Department of Internal Medicine, University of Utah Medical Center, Salt Lake City, UT 84132, [4]Myriad Genetics Inc., 421 Wakara Way, Salt Lake City, UT 84108, [5]Human Genome Center, Building 74, Lawrence Berkeley Laboratory, Berkeley, CA 94720, [6]Division of Biology, 147-75, California Institute of Technology, Pasadena, CA 91125, [7]Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, [8]Department of Preventive Medicine, Creighton University, Omaha, NE, USA and [9]Department of Medical Genetics, McGill University, Montreal, Quebec, Canada

Received August 26, 1994; Revised and Accepted September 28, 1994

BRCA1, a breast and ovarian cancer susceptibility locus, has been isolated and maps to 17q21. A physical map of the _BRCA1_ region which extended from the proximal boundary at D17S776 to the distal boundary at D17S78 was constructed and consists of 51 sequence tagged sites (STSs) from P1 and YAC ends, nine new short-tandem repeat (STR) polymorphic markers, and eight identified genes. The contig, which spans the estimated 2.3 Mb region, contains 29 P1s, 11 YACs, two BACs, and one cosmid. Based on key recombinants in two linked families, _BRCA1_ was further localized to a region bounded by D17S1321 on the proximal side and D17S1325 on the distal side. Within this estimated 600 kb region, the contig was composed completely of P1s and BACs ordered by STS-content mapping and confirmed by DNA restriction fragment fingerprinting.

INTRODUCTION

The isolation of genes through positional cloning requires a minimal region containing the gene of interest, usually defined by identification of key recombinant meiotic breakpoints in families, and a detailed physical map composed of cloned segments of DNA which span the region. A general strategy to clone and map large regions of human DNA using STS-content mapping was demonstrated for the region of chromosome 7 containing the cystic fibrosis gene (1). Huntington's disease (HD) provides an example of the creation and utilization of a detailed physical map to isolate a gene by positional cloning. Initially, physical maps of the defined genetic region were constructed using yeast artificial chromosome (YAC) clones (2,3). The presence of deletions and chimerisms and the large DNA insert size of these vectors made it clear that smaller clones would be more useful for positional cloning. Subsequently, within this mapped region, cosmid contigs (4) and cosmid/P1 contigs (5) were developed which allowed for the identification of several genes in the region, as well as the identification of the HD gene itself (6).

The most recent success of a positional cloning strategy is the isolation of *BRCA1*, a gene for familial ovarian and early-onset breast cancer (7). The search for *BRCA1* began in earnest in 1990 with a report of genetic linkage to a small region on the long arm of chromosome 17 (8). Based on multipoint linkage analysis, *BRCA1* was initially localized to the interval between D17S250 and D17S579 (9). Subsequently, additional recombinants were found in linked kindreds which refined the proximal boundary of the *BRCA1* region to D17S776 (10) and the distal boundary to D17S78 (11). Recently, a physical map of the *BRCA1* region from ERBB2 to D17S509 was published which was comprised primarily of YACs (12).

As part of our effort to isolate the *BRCA1* gene, the strategy was to: 1) create a physical map of the large defined *BRCA1* region; 2) isolate polymorphic short tandem repeat markers (STRs) from these initial physical resources; 3) use these STRs in our set of linked kindreds which contained informative recombination events to refine the region containing *BRCA1*; and then 4) construct a finer and more detailed physical map in this narrower region. This strategy proved to be a key factor in our successful isolation of *BRCA1*.

In this report, we present our preliminary physical map of the initial region from D17S776 to D17S78, detail the genetic events which localized *BRCA1* to a smaller region, and show the P1-based physical map of this smaller region.

RESULTS

In order to develop a physical map of the refined *BRCA1* region as quickly and efficiently as possible, we initiated screens for YACs and P1s at four starting points whose relative locations

*To whom correspondence should be addressed

Genomic clones × **STS markers used in map construction**

STS markers (columns, in order): •686L, •D17S776, •686L, •900SP6, •303R, •749T7, •900T7, •EDH17B2, •749SP6, •D17S1320, •CF4T7, CB5, •1062T7, •28L, 1052SP6, 1051SP6, •1478T7, •1309T7, •D17S1321, •1478SP6, •Ki Antigen, •303L, •1308T7, •1239T7, •682L, •694SP6, 694.5, •D17S1328, •1239SP6, •784SP6, •D17S865, •978SP6, •D17S1322, •784T7, •1141SP6, •694T7, •D17S1323, •978T7, •694.5, •1141T7, •402T7, •283T7, •A167L, •D17S1327, •CA125, •402SP6, •1046T7, 402.5, •406T7, •D17S1326, •1046SP6, 1046.1, •RNU2, •D17S1324, •384SP6, 931SP6, •417SP6, •A167R, •D17S1325, •417T7, 431T7, •380T7, •B194L, •142L, •D17S1329, •380SP6, •D17S858, 241T7, 363T7, •343SP6, •MOX1, TR, 240SP6, •343T7, •489SP6, •VHR, •489T7, 490SP6, •577SP6, •677T7, •604T7, •D17S78

NAME	ADDRESS
*YAC686H7	CEPH
*YAC656A5	CEPH
*YAC714G3	CEPH
*P1-1073	897A1
P1-1074	1033D11
*P1-900	1331B4
*YAC303G8	CEPH
*YAC26F3	CEPH
*YAC682E10	CEPH
*P1-749	904A6
*CF4	COSMID
*P1-1376	355H6
P1-1377	1026F12
P1-1378	1086A1
*P1-1051	1233F3
*P1-1052	1367F12
*P1-1626	949D6
P1-1626	1288A3
P1-1474	762G1
*P1-1478	762G1
*P1-1308	303E3
*P1-1309	355H8
*P1-1238	565D11
P1-1240	1410E5
P1-1237	409H7
P1-1241	1430G4
*P1-1239	603B12
YAC306G8	CEPH
*YAC641A11	CEPH
*BAC694	694B12
*BAC283	283G2
*P1-784	997F1
*P1-978	746B4
*P1-1141	1363C4
*P1-402	445G
*YACA167E6	ST. LOUIS
*P1-1046	102A12
*P1-406	419A
*YACB196A12	ST. LOUIS
*YAC142H4	CEPH
*P1-931	593F8
P1-382	280G
P1-383	311G
*P1-384	1102E
*P1-417	749G
*P1-431	272G
P1-432	372D
*YACB194E6	ST. LOUIS
*P1-380	348E
*P1-241	1466D
P1-240	78E
P1-243	650H
P1-344	1092E
P1-363	782B
*P1-343	9H
*P1-489	201F
P1-490	879A
P1-491	1454A
P1-492	41E
*P1-577	1128C6
P1-578	1477D6
P1-689	337B10
P1-604	1104E1

were known. The map was constructed by utilizing STS markers in the region, isolating clones, and then walking from the ends of the developing contigs using STSs developed from those ends. The STS-content data for all STSs tested against each of the cloned DNA segments is shown in Figure 1. To rapidly merge contigs, we screened initially for YACs as their insert sizes are larger than BACs, P1s, and cosmids. At least 10 of 25 YACs identified in this study appeared to be chimeric or contained deletions, so we independently covered the smallest region containing *BRCA1* in P1s and BACs.

Initially, several unlinked contigs were created around the four starting points in the region; D17S776 (13), EDH17B2 (14), D17S855 (15), and RNU2 (16). The D17S776 and EDH17B2 contigs were oriented and connected with P1-900 and YAC 714G3 (Fig. 2a). This proximal contig was then connected to the D17S855 contig by YAC 652E10 (Fig. 2a). The RNU2 contig was joined to our previous physical map (Miki *et al.*, submitted) at D17S78 through P1 walking. At this point, there were two contigs, a proximal contig containing D17S776, EDH17B2, and D17S855, and a distal contig containing RNU2 and D17S858 (15). The contig was completed with identification of BAC 283 and P1-402 (Fig. 2a).

As the contig was being assembled, walking in a P1 library was done to independently isolate regions covered by YACs and to merge contigs. When P1 screens were negative, cosmid subclones (from YACs) or a BAC library were screened. To close the P1 gaps from 754SP6 to 1239SP6 and from 1141T7 to 402T7, BACs 694B12 and 283G2 were isolated. BAC 694 was shown to contain both of these STSs (Fig. 1). BAC 283 was shown to encompass P1-754, P1-975, and P1-1141 and included part of P1-402 (Fig. 2a).

A total of 25 YACs were identified which mapped to the region. Of these, 11 YACs are included on the map in Figure 2a because they: 1) span areas where there is no other coverage; 2) link smaller contigs; 3) contain STSs used to isolate P1s; or 4) contain STRs which were identified by screening cosmid subclones from those YACs. Fifty-two P1s were identified which mapped to the region, of which 29 are included on the map. Only the minimal set of P1s found which are necessary for a complete contig and/or contain STRs are included on the map.

To aid in construction of the physical map and to further refine the *BRCA1* region, polymorphic STRs were developed by sequencing with CA dinucleotide repeat primers or by direct capture of a series of di-, tri-, and tetranucleotide repeats. STRs D17S1320, D17S1326, D17S1327 (ED2; 10), D17S1325, and D17S1329 were isolated from cosmid subclones and STRs D17S1321, D17S1328, D17S1322, D17S1323, and D17S1324 were from P1s. In Table 1, the primer sequences, size of the product, number of alleles and percent heterozygosity of the new STRs are presented.

In order to define the minimal region containing the *BRCA1* gene, we tested the new STRs in two previously defined recombinant breakpoints. For defining the proximal boundary, kindred 2082 contained the relevant recombination event. Kindred 2082 is the largest *BRCA1*-linked breast/ovarian cancer family studied to date. It has a lod score of 9.5 providing unequivocal evidence for 17q linkage. This family is described in Goldgar *et al.* (10), in which a critical recombinant placing *BRCA1* distal to D17S776 was identified. This recombinant occurred in a woman diagnosed with ovarian cancer at age 45 whose mother had ovarian cancer at age 63. The affected mother was deceased, but could be inferred from her children to have the linked haplotype, which was also present in the 36 other linked cases in the family, in the region between D17S776 and D17S579. Her affected daughter received the linked allele at D17S1328, D17S1327, and more distal loci, but received the allele on the non-BRCA1 chromosome at D17S776, D17S1320 and D17S1321. The recombinant described above therefore placed *BRCA1* distal to D17S1321.

For defining the distal *BRCA1* boundary, kindred 1813 was analyzed. Kindred 1813 is a small family with four cases of breast cancer diagnosed under the age of 40 whose mother had breast cancer diagnosed at age 45 and ovarian cancer at age 61. This family yields a maximum multipoint lod score of 0.60 with 17q markers and given that there is at least one case of ovarian cancer, results in a posterior probability of being a *BRCA1*-linked family of 0.94. This family contains a directly observable recombination event in a woman who developed breast cancer at age 34 which previously indicated that BRCA1 was proximal to D17S78 (11). The genotype of her affected mother at the relevant 17q loci was inferred. The affected individual inherited the *BRCA1*-linked alleles for D17S250, THRA1, D17S800, D17S855, D17S1326, and D17S1324. For markers distal to D17S1324, i.e., D17S1325, D17S1329, D17S858, and D17S78, she inherited the alleles located on the non-disease bearing chromosome. The evidence from this family therefore placed the *BRCA1* locus proximal to the marker D17S1325.

The minimal region containing *BRCA1*, bounded proximally by D17S1321 and distally by D17S1325, was contained within a P1 and BAC contig constructed based on STS content (Fig. 1). In order to confirm the order and continuity of this contig, restriction fragment fingerprinting was performed. Both *Hin*dIII and *Eco*RI digests among P1s and BACs in the region were compared for overlaps among the clones. The *Eco*RI restriction fragment fingerprints are shown in Figure 3. Further evidence for the overlap of clones came from observations of identical DNA sequences on adjacent P1s (data not presented).

With refinement of the physical map, the P1s and BACs were used as probes, selectors, and templates for direct screening of cDNAs, hybrid selection of cDNAs, and random sequencing, respectively (7). The *BRCA1* gene (7) is located on BAC 283, totally encompasses P1-754, and contains parts of BAC 694, P1-975, and P1-1141, as expected due to the overlapping nature of these clones (Fig. 2a). STRs D17S1323, D17S855, and D17S1322 are located in introns of the *BRCA1* gene. The exact size of the gene is not yet known, but is estimated to be 100 kb.

Other identified genes were localized with respect to our physical map. EDH17B2 (14) was used as one of the starting

Figure 1. P1, YAC, BAC, and cosmid clones in the *BRCA1* region. The clones, STS markers and genes used to construct Figure 2a are preceded by a*. Markers that are present in a clone are indicated with a + and those that are absent are indicated with a −. Where there is no indicator, the STS was not tested against the clone. YAC origin and individual plate and well numbers for the P1s and BACs are listed under the Address column.

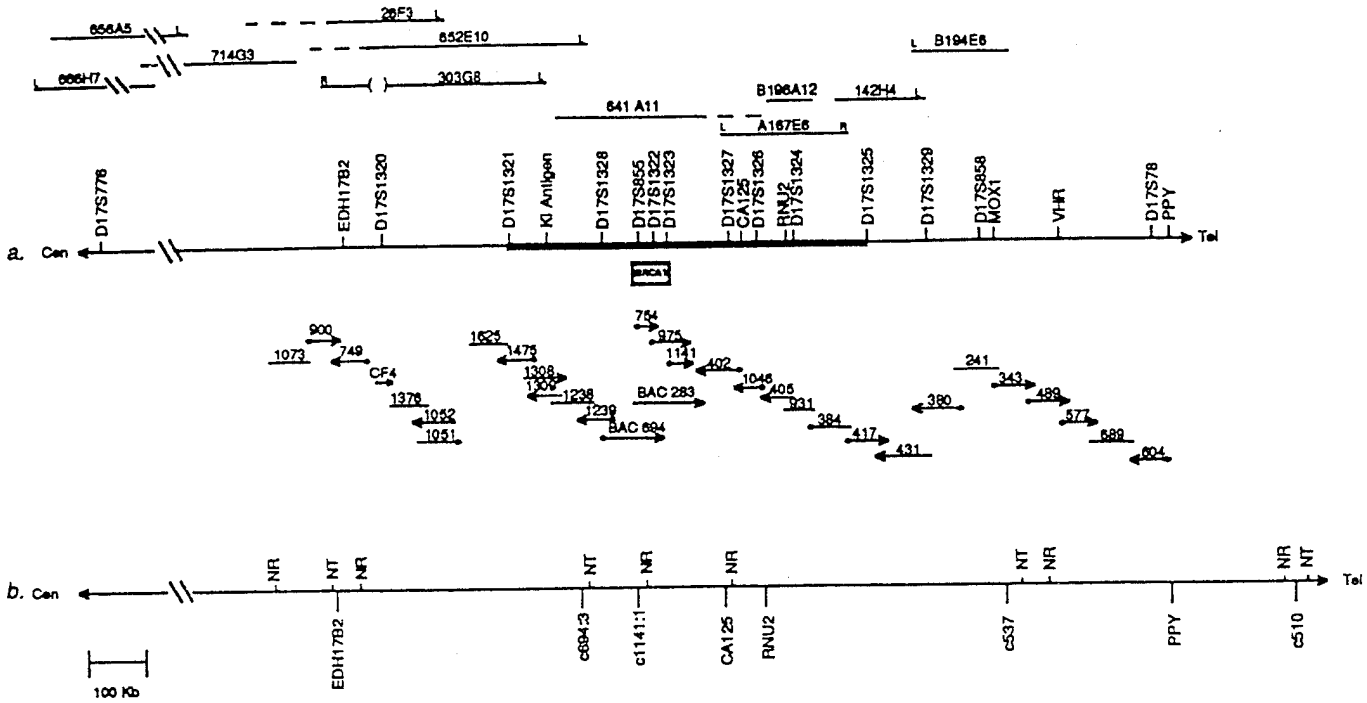


Figure 2. Physical maps of the *BRCA1* region. The two maps are drawn to scale and overlap in the region between EDH17B2 and PPY. Cen and tel denote toward the centromere and telomere, respectively. (a) Physical map of YACs, P1s, BACs, and cosmid clones from D17S776 to PPY. Only the minimal sets of P1 clones and YACs for the contig are shown. The refined *BRCA1* region is contained within the bold line of the map. The location of the *BRCA1* gene is shown in the box and is described in the text. Polymorphic markers and genes are indicated by vertical lines on the map. YAC clones are drawn above the map. L and R represent the left and right arms of the YAC vector and indicate the STS developed for walking. Dashed lines indicate the chimeric portions and the parenthesis indicates a deletion within the YAC. P1 clones, BACs 283 and 694, and cosmid CF4 are drawn below the map. The T7 primer end of the P1 clone insert is indicated by an arrow on the solid line. The other end of the P1 clone insert is flanked by the SP6 primer, and that of the cosmid insert is flanked by the T3 primer which are shown as dots. The arrow and dot indicate that a STS was developed for walking. (b) Pulsed field gel restriction map for a 2 Mb region containing EDH17B2 and D17S78. Markers are localized to a restriction fragment; exact locations are unknown. NT and NR represent *Not*I and *Nru*I sites that are unmethylated in DNA from MRC5 normal human fibroblasts and human mammary epithelial cells. Fibroblast and mammary epithelial cell DNA yielded similar results.

Table 1. New STR markers located in the region between EDH17B2 and D17S78

Locus	Primer name	Primer sequences	Approx size STS product	No of alleles	Het[a]	*n*
D17S1320	CF4	ACT TTC CAG AAA ATC TCT GCT C CCA CGT CTT TTC TGT GTT CC	180	6	0.63	70
D17S1321	tdj1475	CAG AGT GAG ACC TTG TCT CAA A TTC TGC AAA CAC CTT AAA CTC AG	170	13	0.84	62
D17S1328	tdj1239	TCT GTA GGT GCT AAG CAG TGG GTT GCA GTG AGC CGA GAA TG	250	10	0.50	68
D17S1322	s754	CTA GCC TGG GCA ACA AAC GA GCA GGA AGC AGG AAT GGA AC	130	7	0.67	72
D17S1323	s975	TAG GAG ATG GAT TAT TGG TG AAG CAA CTT TGC AAT GAG TG	155	6	0.44	68
D17S1326	AA1	CAG CTG ATA TTT CAC AGG ACT AGA GCA AAA CTC CAT CTC AAA CA	90	10	0.83	40
D17S1324	cdj931	GTG GAG TGG GTG GCG ACC T CAC ACG TGT ACC TAT GTT CC	340	7	0.72	50
D17S1325	U5	AAA GGT GGC AAT TCA CAG TTG GTG ATA AAA CTC AGT GGT ACT C	155	13	0.85	68
D17S1329	VRS31	GAC TCT GAA GGT AAA GAG CAA CTC CCC TGC CTT GGG AGT AG	170	7	0.76	42

[a]Observed heterozygosity from *n* independent chromosomes.

points for the map. PPY (17) was localized to within 45 kb of D17S78, based on data that a cosmid contained both of them (18). Other genes mapped by STSs were Ki antigen (19), human L21 (12), MOX1 (20), VHR (21), CA125 (22), and RNU2 (16). RNU2 is a tandem array of genes of approximately 120 kb (16) and is localized on the map based on a STS that mapped to P1s 405 and 931 (Fig. 1).

The size of the region from D17S776 to D17S78 is estimated to be at least 2.3 Mb based on the sum of size of YAC 714G3 and the pulsed field map. The distance from D17S776 to EDH17B2 is a minimum of 800 kb based on the size of YAC 714G3 (12) which lies between them. The size of the region from EDH17B2 to D17S78 is approximately 1.5 Mb based on a pulsed field map constructed from *Not*I and *Nru*I digests (Fig. 2b). For the region from D17S1321 to D17S1325, estimates of sizes of the P1s and BACs and degree of overlap among clones were computed based on restriction fragment fingerprints (Fig. 3, Table 2). The distance from D17S1321 to D17S1325 is estimated to be no more than 600 kb. This estimate is consistent with the pulsed field mapping results.

DISCUSSION

This report presents a physical map of the minimal *BRCA1* region completely subcloned into bacterial vectors. It includes all the P1s, BACs, and STSs comprising the minimal tiling path. Nine polymorphic STRs were identified which were used to refine the region genetically. These nine STRs and the 51 end-clone STSs were used to screen for P1s and to better determine P1 and BAC overlaps. We determined that the size of the region between D17S1321 and D17S1325 was less than 600 kb, within which lies the recently isolated *BRCA1* gene (7).

Contig assembly began with YACs and proceeded to P1s, BACs, and cosmid subclones which are more convenient reagents for gene isolation and are less prone to deletion and chimerism. Albertsen *et al.* (12), during construction of their *BRCA1* physical map, found that several parts of the map were not represented by YACs due to small deletions or rearrangements. They were able to overcome most of the problems identified by isolating P1s. We also found that many of the YACs were chimeric or contained deletions. Of the YACs in Figure 2a, the right arms of 26F3 and 652E10 were chimeric, 303G8 contained an internal deletion, and the distal portion of 641A11 was chimeric. Other YACs such as 712A7 and 693E8 were not included on the map due to a three-part chimera where neither arm was on chromosome 17, nor was 851H11 included due to a large deletion. P1s and BACs are reported to be more stable and less prone to deletion and chimerism (23,24). In addition, it is relatively easy to isolate microgram quantities of insert DNA from BACs and P1s. For the 600 kb map of the refined *BRCA1* region reported here, there is complete coverage with P1s and BACs, with no known gaps of DNA.

The size of the region from D17S776 to D17S78 is estimated to be 2.3 Mb, with the distance from EDH17B2 to PPY estimated to be 1.5 Mb based on long-range pulsed field restriction mapping. Within the refined *BRCA1* region defined by D17S1321 and D17S1325, the distance is estimated to be approximately 600 kb based on restriction fragments of the P1s and BACs, which is consistent with the pulsed field gel results. Slight differences between the two maps are likely due to experimental variance in measurements. In the pulsed field analysis, we only can determine the fragment on which a particular marker is located. The exact location of markers on large restriction fragments cannot be ascertained. For the size estimate based on *Eco*RI and *Hin*dIII restriction fragments, distances are imprecise for insert/vector junction fragments where the two P1s overlap, as they will not share a common restriction fragment. In addition,

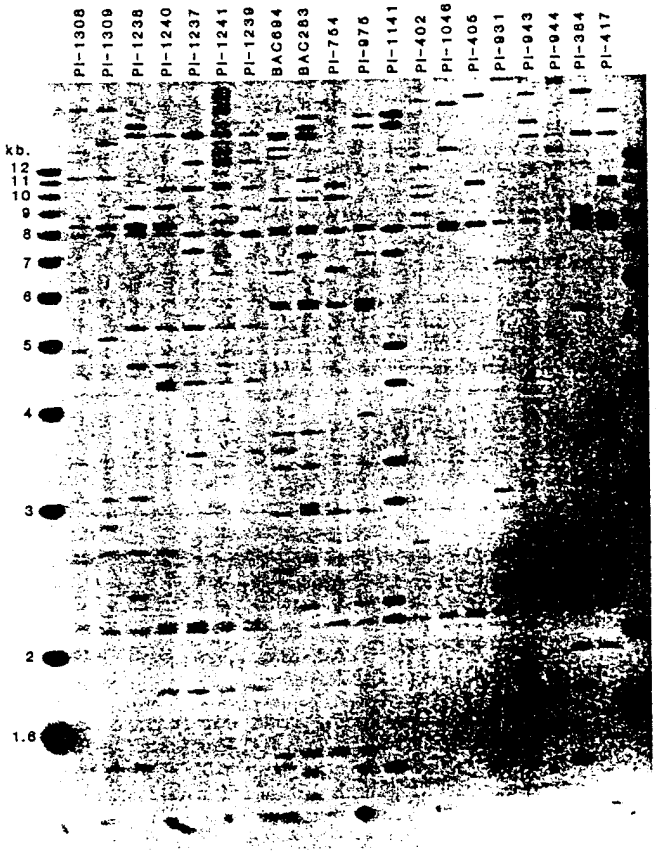


Figure 3. Restriction fragments generated with *Eco*RI from P1s and BACs between D17S1321 and D17S1325. Not all the P1s shown here were included in Figure 2a.

Table 2. Estimated sizes for P1s and BACs between D17S1321 and D17S1325

Clone	Size (Kb)
P1-1309	70
P1-1308	85
P1-1238	70
P1-1239	65
BAC 694	110
BAC 283	130
P1-754	40
P1-975	83
P1-1141	60
P1-402	76
P1-1046	41
P1-405	35
P1-931	43
P1-384	70
P1-417	78

we did not investigate the RNU2 region in detail, and there may be small deletions of the P1s within the RNU2 gene cluster based on the smaller insert sizes of P1-405 and P1-931 which are 35 and 43 kb respectively versus an average P1 size of 80 kb (25).

These clones facilitated systematic screening of cDNA libraries for transcribed sequences, followed by mutation screening of candidate genes in *BRCA1* carriers. Although the purpose for which the map was originally constructed (i.e., isolation of *BRCA1*) has now been completed, the physical map and STR information presented in this manuscript are still of value for researchers studying the breast cancer gene and other genes in the region. Over the past several years, the *BRCA1* region has been intensively studied by many research groups. With a collaborative effort to compile this wealth of physical mapping and DNA sequence data, it should be possible to investigate the complete genomic structure of this gene-rich segment of human DNA.

MATERIALS AND METHODS

Vectors

YACs. The YAC libraries used were from the Centre d'Etude du Polymorphisme Humain (CEPH YAC library) (26) and Washington University (St. Louis YAC library) (27). YACs used in the construction of the map are listed in Figure 1. Positive screenings of YAC 26F3 with EDH17B2 and YACs B196A12 and A167E6 with RNU2 were reported by F.Collins at a Chromosome 17 Workshop (28). YAC 303G8 was positive with EDH17B2 as reported to GDB by S.Thomas (29).

For screening libraries, PCR reactions were carried out in a 10 μl volume, containing 1 ng pooled YAC DNAs, 0.5 μM each primer, 200 μM dNTP, 10 mM Tris−HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, and 0.25 U of *Taq* polymerase in a Thermal Cycler (GeneAmp PCR System 9600, Perkin-Elmer Cetus) for 35 cycles of 94°C for 10 s, 55°C for 10 s, 72°C for 10 s. Products were visualized on 1.2% agarose gels stained with ethidium bromide.

YAC vector-*Alu* PCR was performed to isolate YAC end sequences. Primers used for PCR were inter-*Alu* primers 1IL, 2IL, 2IR, and 2MR and PFYL for the left arm of the YAC vector and PFYR for the right arm of the YAC vector (Table 3). PCR conditions were the same as described above. The PCR products from the primer pair of *Alu* primer and YAC vector primer were compared to products using only *Alu* primer. Unique *Alu*-vector products were isolated and sequenced. When isolation of YAC end sequences by YAC vector-*Alu* PCR was unsuccessful, the same procedure was done using a set of 12 random primers instead of the *Alu* primers. The PCR products from the primer pair of a random primer and YAC vector primer were compared to products using only the random primer. Unique random primer-vector products were sequenced. This allowed us to sequence the ends of all the YACs. All end-YAC STSs were checked against somatic cell hybrids, ME3-5 and ME3-7 (Cerosaletti *et al.*, submitted) to ensure that they were on 17q12−q21. DNA sequencing was done by the chain termination method (30) using Sequenase (USB, Cleveland, OH) or by cycle sequencing using a Cyclist™ *Taq* DNA Sequencing Kit (Stratagene, CA).

P1s. The P1 library (DMPC-HFF # 1; 25) was screened by a PCR-based strategy and positive clones were isolated by GENOME SYSTEMS, INC. (Missouri). P1s and BACs reported here are available for research purposes at no cost upon request to the corresponding author. P1 DNAs were prepared using an alkaline lysis plasmid isolation protocol and purified by equilibrium centrifugation in CsCl−ethidium bromide gradients. P1 inserts were sequenced with T7 and SP6 as described above for YAC sequencing. End-specific STSs were used for the next screening of the P1 library.

BACs. BAC screening was by a hybridization-based strategy with STS 754SP6. A radioactive probe was annealed with total human DNA before adding to the hybridization buffer. Forty-two filters containing 96,000 BAC clones were hybridized and washed twice with 2×SSC−0.1% SDS at 47°C. Positive clones were identified and confirmed by PCR using the 754SP6 STS.

Cosmids. Cosmid libraries were constructed from each of the YACs. Yeast strains containing YACs were grown in AHC medium (1.7 g/l yeast nitrogen base (Difco), 5 g/l ammonium sulfate, 10 g/l casein hydrolysate, 20 g/l glucose, 20 mg/l adenine hemisulfate; pH 5.8 (27)). High molecular weight DNAs were prepared from YAC-containing yeast strains by centrifugation through a sucrose-step gradient using the methodology of Carle and Olson (31). Cosmid libraries were constructed from total yeast DNAs as described by Tokino (32). The resultant cosmid libraries were plated at a density of 15−20 clones/cm^2, and clones containing human DNA were identified by colony hybridization using human Cot-1 DNA (Life Technologies, Inc. MD) as a probe. Selected cosmids were sequenced with T3 and T7 primers (Table 3) to make end-specific STSs of cosmid inserts.

Genotyping methods

Genotyping of family members for the markers described were carried out at the University of Utah and McGill University for kindreds 2082 and 1813, respectively. In both locations, similar, standard PCR protocols and conditions were used with the exception that ^{35}S-radiolabelling was used at McGill, while ^{32}P-labelling was performed at Utah. All autoradiographs at both locations were read by two independent observers and all events critical to recombinant analysis were repeated to verify the initial results.

Identification of STRs

STR identification by sequencing. To identify CA repeats in the contig, cosmids were screened by hybridization to a radiolabelled $(CA)_{10}$ oligonucleotide probe. Both positive P1 and cosmid clones containing human inserts were sequenced directly with two sets of three mixed primers: CA1 and CA2 (Table 3). The first

Table 3. Primers and sequences

Name	Primer Sequences	Description
1IL	GGG CTG CAG CAT GTT GGC CAG GCT GGT CT	Inter-*Alu*
2IL	GGA GTG CAG TGG CGC GAT CTC TGC AGA CTG	Inter-*Alu*
2IR	CAG TCT GCA GAG ATC GCG CCA CGC ACT GCA GCC TGG	Inter-*Alu*
2MR	GCA CTG CAG CCT GGG CAA CAG AGC GAG	Inter-*Alu*
PFYL	CAA GTT GGT TTA AGG CGC AAG	YAC left arm
PFYR	TCG AAC GCC CGA TCT CAA G	YAC right arm
YL	TGG AAG GAC CGG ATA ATT ATT TGG AAT CTC	YAC left arm
YR	GTA GAG CTT GAA AGA AAT ATT TCA ACT TGC	YAC right arm
T3	ATT AAC CCT CAC TAA AG	Cosmid
T7	AAT ACG ACT CAC TAT AG	Cosmid, P1, BAC
SP6	GAT TTA GGT GAC ACT ATA G	P1, BAC
CA1	ACA CAC ACA CAC ACA CAC N (N is T or C or G)	
CA2	CAC ACA CAC ACA CAC ACA N	
RP-2	TGA GTA GAA TTC TAA CGG CCG TCA TTG TTC	
RL.blunt	(PO_4) GAA CAA TGA CGG CCG TTA GAA TTC TAC TCA (NH_2)	
RL.S3A	(PO_4) GAT CGA ACA ATG ACG GCC GTT AGA ATT CTA CTC A	
RP-B	(PO_4) TGA GTA GAA TTC TAA CGG CCG TCA TTG	

Oligonucleotide sequences for the primers are listed in the 5′−3′ direction.
RP-2.blunt is RP-2 annealed with RL.blunt. RP-2.S3A is RP-2 annealed with RL.S3A.

STR primer for each marker was designed from this sequence and used to sequence back through the CA repeat. The second primer was designed so that the marker would generate a PCR product of approximately 150 bp. Both primers were designed so that the annealing temperature was 55°C. This primer pair was then screened against chromosome 17 somatic cell hybrids to ensure that it mapped to chromosome 17.

STR cloning by direct capture. For each P1, 1.5 μg DNA was digested independently with *Sau*3AI and *Hae*III. 0.5 μg of each digest was examined on an agarose gel; the remaining DNA was phenol–chloroform extracted, isopropanol precipitated, and redissolved at a concentration of 50 ng/μl. The double stranded anchor adapter RP-2.blunt (Table 3) was ligated to the *Hae*III digested P1 DNA. Similarly, the double stranded anchor adapter RP-2.S3A (Table 3) was ligated to the *Sau*IIIA digested P1 DNA. The two ligation reactions from each P1 were combined and fractionated over a sepharose CL-6B column to remove excess anchor adapters. Approximately 25 ng of anchored P1 DNA was hybridized in 50 μl of 20 mM Hepes pH 7.5, 1 mM EDTA, 3 M tetramethylammonium chloride, for 60 min at 50°C with mixtures of 23mer biotinylated microsatellite repeat oligonucleotides (each oligo at 0.1 nM) (e.g. biotin-CA: BTB CAC ACA CAC ACA CAC ACA CA where B = biotin amidite). Oligos for CA, GA, AAC, AAG, AAT, ACC, AGC, AGG, TCA, CGC, AAAC , AAAG, AAAT, AAGG, and AGGG were used in this study. P1 DNA fragments hybridized to the biotinylated oligos were captured on streptavidin paramagnetic particles (Dyna), resuspended in 100 μl of 20 mM Hepes pH 7.5, 1 mM EDTA, 3 M tetramethylammonium chloride, and incubated for 10 min at 55°C. The biotinylated oligos and hybridized P1 DNAs were then captured and washed twice with 100 μl of 10 mM Tris pH 8.2, 50 mM KCl, 0.1 mM EDTA, and resuspended in 10 μl of 5 mM Tris pH 8.2, 0.1 mM EDTA. 1.5 μl aliquots of the resuspended beads and captured DNAs were PCR amplified using the primer RP-B. The PCR reactions were run out on agarose gels. Microsatellite repeat bearing bands were cut out of the gels, purified on Qiaex beads (Qiagen), ligated into pUC 18, cloned, and sequenced.

Restriction mapping of P1 and BAC clones

DNAs were isolated from a 1.5 ml overnight culture of each strain (33). The DNA was digested with *Eco*RI and *Hin*dIII separately, labelled with ^{32}P-dATP by filling-in with the Klenow fragment of *E.coli* DNA polymerase I, and separated in a 0.6% agarose gel by electrophoresis. The gel was then dried onto DE81 filter paper under vacuum and exposed to Kodak X-ray film. The computer program *GEL* by J.-M.Lacroix was used to assign fragment sizes. Size was determined by summing the restriction fragments, excluding the vector fragments. Overlap between P1s and BACs was calculated by summing the restriction fragments in common.

Pulsed field gel restriction mapping

MRC5 normal human fibroblasts and human mammary epithelial cells were embedded in 1% low melt agarose at a concentration of 2.5×10^7 cells/ml using gel syringes (NEB). Hardened agarose rods were extruded into 10 ml of digestion solution consisting of 1.0 mg/ml proteinase K (Boehringer), 100 mM EDTA, and 1% *N*-laurylsarcosine (Sigma). Rods were incubated at 50°C for 48 h at which time they were chilled on ice and dialyzed three times in 50 mM EDTA in a volume of 50 ml/wash. Plugs were stored in the gel syringes at 4°C. Twenty microliter slices of the embedded DNA were cut from the syringes and dialyzed in TE for 1 h on ice. This solution was removed and replaced with 200 μl of the appropriate 1× restriction buffer and dialyzed on ice for 30 min. This solution was replaced with 1× restriction buffer containing 40 to 100 units of *Not*I or *Nru*I and further incubated for 30 min on ice. Digestions were at 37°C for 6 h. The restriction buffer was removed and replaced with 1 ml of ice cold gel running buffer (0.5×TBE).

Gels were run on either a Pulsphor (LKB) or CHEF DRII unit (BioRad). Agarose gels (1%) in 0.5×TBE were run in 0.5×TBE at a temperature of 14°C. Running conditions varied with fragment size being resolved, but consisted of continuously ramped pulses running at 200 V. Markers consisting of yeast chromosomes or lambda DNA were purchased from NEB. Gels were stained in 0.5 μg/ml EtBr for 30 min and destained in distilled water for 30 min. Gels were photographed and nicked in a UV Stratalinker 2400 (Stratagene) on the auto-crosslink setting. Gels were then placed in alkali (0.5 M NaOH, 1.5 M NaCl) for 30 min, transferred via capillary blotting onto GeneScreen Plus (DuPont) in the same solution for 24 h, and neutralized in 0.5 M Tris/3.0 M NaCl. Filters were hybridized to radiolabelled probes in Hybrisol II buffer (Oncor), washed to 2×SSC/1%SDS at 65°C, and exposed to film at −70°C for 1–4 days. EDH17B2, RNU2, and PPY probes were generated by PCR with primers described in the Chromosome 17 Workshop Report (34). The remaining probes used to construct this map were cDNAs isolated by solution hybrid capture (20) during the course of generating a transcript map of the *BRCA1* region (P.A.Futreal *et al.*, in preparation). The map (Fig. 2b) was constructed by aligning films of consecutive probings and measuring fragment sizes.

Electronic information access

DNA primer sequences and sizes of the products are available from GDB. This information may also be obtained directly from the authors via anonymous ftp to morgan.med. utah. edu (128.110.62.50) in the directory pub/STS.

ACKNOWLEDGEMENTS

We thank K.Fournier from Fred Hutchinson Cancer Research Center for supplying us with the somatic cell lines. We thank M.Ashton for technical assistance in drawing the figures and L.M.Bennett for confirmation of STS-content. We are grateful to D.Smoller and K.Schriewer from Genome Systems for their careful and efficient screening of the P1 library. This work was supported by grants CA-48711 and CA-55914 from the National Institutes of Health, by Myriad Genetics, the National Cancer Institute of Canada Breast Cancer Initiative, the Canadian Genetics Diseases Network, and the Cedars Cancer Institute of the Royal Victoria Hospital.

REFERENCES

1. Green, E.D. and Olson, M.V. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* 250: 94–98.
2. Zuo, J., Robbins, C., Taillon-Miller, P., Cox, D.R., and Myers, R.M. (1992) Cloning of the Huntington disease region in yeast artificial chromosomes. *Hum. Mol. Genet.* 1: 149–159.
3. Bates, G.P., Valdes, J., Hummerich, H., Baxendale, S., LePaslier, D.L., Monaco, A.P., Tagle, D., MacDonald, M.E., Altherr, M., Ross, M., Brownstein, B.H., Bentley, D., Wasmuth, J.J., Gusella, J.F., Cohen, D., Collins, F., and Lehrach, H. (1992) Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region. *Nature Genet.* 1: 180–187.
4. Zuo, J., Robbins, C., Baharloo, S., Cox, D.R., and Myers, R.M. (1993) Construction of cosmid contigs and high-resolution restriction mapping of the Huntington disease region of human chromosome 4. *Hum. Mol. Genet.* 2: 889–899.
5. Baxendale, S., MacDonald, M.E., Mott, R., Francis, F., Lin, C., Kirby, S.F., James, M., Zehetner, G., Hummerich, H., Valdes, J., Collins, F.S., Deaven, L.J., Gusella, J.F., Lehrach, H., and Bates. G.P. (1993) A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene. *Nature Genet.* 4: 181–186.
6. The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72: 971–983.
7. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., Skolnick, M.H. (1994) Isolation of a strong candidate for the 17q-linked breast and ovarian cancer susceptibility gene. BRCA1. *Science* 266: 66–71.
8. Hall, J.M., Lee, M.K., Morrow, J., Newman, B., Morrow, J., Anderson, L., Huey, B., King, M.-C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250: 1684–1689.
9. Easton, D.F., Bishop, D.T., Ford, D., Crockford, G.P., and the Breast Cancer Linkage Consortium. (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am. J. Hum. Genet.* 52: 678–701.
10. Goldgar, D.E., Fields, P., Lewis, C.M., Tran, T.D., Cannon-Albright, L.A., Ward, J.H., Swensen, J., Skolnick, M.H. (1994) A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis. *J. Natl Cancer Inst.* 86: 200–209.
11. Simard, J., Feunteun, J., Lenoir, G., Tonin, P., Normand, T., The, V.L., Vivier, A., Lasko, D., Morgan, K., Rouleau, G., Lynch, H., Labrie,F., and Narod, S. (1993) Genetic mapping of the breast–ovarian cancer syndrome to a small interval on chromosome 17q12–21; exclusion of candidate genes EDH17B2 and RARA. *Hum. Mol. Genet.* 2: 1193–1199.

12. Albertsen, H.M., Smith, S.A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cross, C.S., Slijepcevic, P., Carlson, M.N., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., MeiSheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B.A.J., and White, R. (1994) A physical map and candidate genes in the BRCA1 region on chromosome 17q12–21. *Nature Genet.* 7: 472–479.
13. Weber, J.L.; Chromosome 1,2,3,4,5,7,11,13,15,16,17, 18 probes; 1992. Personal Communication in GDB(TM) Human Genome Data Base. Id: G00-032-638.
14. Friedman, L.S., Lynch, E.D., and King, M-C. (1993) Two independent polymorphisms at the 17β-hydroxysteroid dehydrogenase (EDH17B) gene (17q21). *Hum. Mol. Genet.* 2: 821.
15. Anderson, L.A., Friedman, L., Osborne-Lawrence, S., Lynch, E., Weissenbach, J., Bowcock, A., and King, M.-C. (1993). High-density genetic map of the BRCA1 region of chromosome 17q12–q21. *Genomics* 17: 618–623.
16. Westin, G., Zabielski, J., Hammarstroem, K., Monstein, H.-J., Bark, C., and Pettersson, U. (1984) Clustered genes for human U2 RNA. *Proc. Natl Acad. Sci. USA* 81: 3811–3815.
17. Takeuchi, T., Gumucio, D., Yamada, D.L., Meisler, M.N., Minth, C.D., Dixon, J.E., Eddy, R.E., and Shows, T.B. (1986) Genes encoding pancreatic polypeptide and neuropeptidase Y are on human chromosome 17 and 7. *J. Clin. Invest.* 77: 1038–1041.
18. Chandrasekharappa, S.C., Friedman, L., King, S.E., Lee, Y.-H., Welsch, P., Bowcock, A.M., Weber, B.L., King, M.-C., and Collins, F.S. (1994) The gene for pancreatic polypeptide (PPY) and the anonymous marker D17S78 are within 45 kb of each other on chromosome 17q21. *Genomics* 21: 458–460.
19. Nikaido, T., Shimada, K., Shibata, M., Hata, M., Sakamoto, M., Takasaki, Y., Sato, C., Takahashi, T., and Nishida, Y. (1990) Cloning and nucleotide sequence of cDNA for Ki antigen, a highly conserved nuclear protein detected with sera from patients with systemic lupus erythematosus. *Clin. Exp. Immunol.* 79: 209–214.
20. Futreal, P.A., Cochran, C., Rosenthal, J., Miki, Y., Swensen, J., Hobbs, M., Bennett, L.M., Haugen-Strano, A., Marks, J., Barrett, J.C., Tavtigian, S.V., Shattuck-Eidens, D., Kamb, A., Skolnick, M., and Wiseman, R.W. (1994) Isolation of a diverged homeobox gene, MOX1, from the BRCA1 region on 17q21 by solution hybrid capture. *Hum. Mol. Genet.* 3: 1359–1364.
21. Kamb, A., Futreal, P.A., Rosenthal, J., Cochran, C., Harshman, K.D., Liu, Q., Phelps, R.S., Tavigian, S.V., Tran, T., Hussey, C., Bell, R., Miki, Y., Swensen, J., Hobbs, M., Marks, J., Bennett, L., Barrett, J.C., Wiseman, R.W., and Shattuck-Eidens, D. (1994) Localization of the VHR phosphatase gene and its analysis as a candidate for BRCA1. *Genomics*: 23: 163–167.
22. Campbell, I.G., Nicolai, H.M., Foulkes, W.D., Senger, G., Stamp, G.W., Allan, G., Boyer, C., Jones, K., Bast, R.C., Solomon, E., Trowsdale, J., and Black, D.M. (1994) A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. *Hum. Mol. Genet.* 3: 589–594.
23. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* 89: 8794–8797.
24. Smoller, D.A., Petrov, D., and Hartl, D.L. (1991) Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75–100 kilobase pairs. *Chromosoma* 100: 487–494.
25. Shepherd, N.S., Pfrogner, B.D., Coulby, J.N., Ackerman, S.L., Vaidyanathan, G., Sauer, R.H., Balkenhol, and Sternberg, N. (1994) Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl Acad. Sci. USA* 91: 2629–2633.
26. Albertsen, H. M., Abderrhim, H., Cann, H. M., Dausset, J., LePaslier, D., and Cohen, D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc Natl. Acad. Sci. USA* 87: 4256–4260.
27. Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D., and Olson, M. V. (1989) Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* 244: 1348–1351.
28. Fain, P. (1992) Third international workshop on human chromosome 17 mapping. *Cytogenet Cell Genet.* 60: 178–186.
29. Thomas, S; Chromosome 17 YACs; 1993. Personal Communication in GDB(TM) Human Genome Data Base. GDB Id: G00-042-978.
30. Sanger, F., Nicken, S., and Coulson, A.R. (1977) DNA sequencing with chain termination inhibitors. *Proc. Natl Acad. Sci. USA* 74: 5463–5467.
31. Carle, G. F., and Olsen, M. V. (1984) Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Res.* 12: 5647–5664.
32. Tokino, T., Takahashi, E., Moro, M., Tanigami, A., Glaser, T., Park, J. W., Jones, C., Hori, T., and Nakamura, Y. (1991) Isolation and mapping of 62 new RFLP markers on human chromosome 11. *Am. J. Hum. Genet.* 48: 258–264.
33. Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
34. Fain, P.R., Solomon, E., Ledbetter, D.H. (1991) Second International Workshop on Human Chromosome 17. *Cytogenet. Cell Genet.* 57: 65–77.

GENOMICS 28, 530–542 (1995)

Generation of a Transcription Map at the HSD17B Locus Centromeric to *BRCA1* at 17q21

JOHANNA M. ROMMENS,*,[1] FRANCINE DUROCHER,† JODI MCARTHUR,* PATRICIA TONIN,‡ JEAN-FRANÇOIS LEBLANC,† TODD ALLEN,* CAROLLE SAMSON,† LORENZO FERRI,§ STEVEN NAROD,‡,¶ KENNETH MORGAN,‡ AND JACQUES SIMARD†

*Department of Genetics, Research Institute, The Hospital for Sick Children, Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, M5G 1X8; †Medical Research Council Group in Molecular Endocrinology, CHUL Research Center and Laval University, Quebec City, Quebec, G1V 4G2; and ‡Departments of Human Genetics and Medicine, McGill University, Montreal General Hospital Research Institute, §Division of Medical Genetics, Royal Victoria Hospital, and ¶Faculty of Medicine, Montreal General Hospital and McGill University, Montreal, Quebec, Canada H3G 1A4

Received February 14, 1995; accepted May 19, 1995

A detailed transcription map of the 320-kb region containing the HSD17B locus on chromosome 17 was generated. Thirty unique cDNA fragments, retrieved following the hybridization of immobilized YACs to primary pools of cDNAs prepared from RNA of mammary gland, ovary, placenta, and the Caco-2 cell line, were aligned into 10 transcription units by physical mapping and hybridization to RNAs of a series of tissues. The cDNAs were then further characterized by sequencing and used to screen mammary gland cDNA libraries. Fragments corresponding to the broadly expressed γ-tubulin and Ki antigen genes were identified. A full-length cDNA clone encoding a 117-amino-acid protein homologous to the rat ribosomal protein L34 was isolated. Portions of genes with restricted patterns of expression were also obtained, including the previously characterized *HSD17B1*. One new gene, for which a full-length cDNA was isolated, was found to have an interesting tissue-specific pattern of expression with abundant mRNA in both the colon and the testis and in the mammary carcinoma cell line BT-474. This contrasted with the barely detectable level observed in several tissues including normal mammary gland. Of the five additional transcription units identified, one showed no similarity, two showed identity to human expressed sequences, and two displayed similarity to genes of animal species by amino acid alignment. These latter cDNA clones include potential homologues of a rat nuclear tyrosine phosphatase and of a factor of *Drosophila* that is known to be involved in the negative regulation of transcription of segment identity genes.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. L38932–941 and L39000.

[1] To whom correspondence should be addressed at the Department of Genetics, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. Telephone: (416) 813-7095. Fax: (416) 813-4931.

INTRODUCTION

The refinement of the breast cancer susceptibility locus within 17q12–q21 by genetic linkage (Hall *et al.*, 1990; Narod *et al.*, 1991; Easton *et al.*, 1993) and recombination mapping studies (Anderson *et al.*, 1993; Bowcock *et al.*, 1993; Chamberlain *et al.*, 1993; Kelsell *et al.*, 1993; Simard *et al.*, 1993; Albertson *et al.*, 1994a; Goldgar *et al.*, 1994; Smith *et al.*, 1994; Tonin *et al.*, 1994) has recently led to the identification of the *BRCA1* gene at D17S855 (Miki *et al.*, 1994; Futreal *et al.*, 1994a). Germline mutations have been detected in linked families in this putative transcription factor (Futreal *et al.*, 1994a; Castilla *et al.*, 1994; Simard *et al.*, 1994; Friedman *et al.*, 1994a). Noted features of these families include premenopausal onset of breast cancer and high incidences of ovarian cancer. In sporadic cases, the involvement of additional genes on the long arm is suspected (Vogelstein and Kinzler, 1994), as studies performed directly on breast tumors indicate that a range of chromosome 17 abnormalities and rearrangements occur in a significant proportion of patients (Sato *et al.*, 1991; Kallioniemi *et al.*, 1994). This has been supported by the high frequency of loss of heterozygosity in the BRCA1 region (Eccles *et al.*, 1990; Futreal *et al.*, 1992; Smith *et al.*, 1992; Saito *et al.*, 1993; Cropp *et al.*, 1994; Neuhausen and Marshall, 1994). These changes could involve genes that are not necessarily causal, but may be involved in the progression of breast cancers. Additional and refined mapping of 17q and delineation of the minimal intervals where alterations occur are needed.

High-density maps of physical markers and a number of candidate genes have been obtained from 17q12–

0888-7543/95 $12.00

q21 (Albertson *et al.*, 1994b; Campbell *et al.*, 1994; Futreal *et al.*, 1994b; Kamb *et al.*, 1994; Neuhausen *et al.*, 1994). A detailed transcription map of the region just telomeric of *BRCA1* (Jones *et al.*, 1994) and preliminary data from a broader map have recently been reported (Friedman *et al.*, 1994b). The HSD17B locus is centromeric of *BRCA1* and was known to include at least two genes, one of which is likely to be a processed pseudogene (Luu-The *et al.*, 1989, 1990). The active gene, *HSD17B1* (formerly known as *EDH17B2*), catalyzes the interconversion of the weak estrogen, estrone, to the more potent 17β estradiol. Its known biological role in the regulation of estrogen, the physical location, and the expression pattern led initially to consideration of this gene as a *BRCA1* candidate; however, affected individuals from families that showed chromosome 17 linkage to breast cancer did not reveal sequence alterations (Simard *et al.*, 1993; Kelsell *et al.*, 1993).

We sought to develop a detailed transcription map with YACs from the locus. The procedures used to identify cDNAs involved hybridization of mixtures of cDNA libraries to immobilized genomic DNA to permit the rapid identification of genes by a modified direct selection procedure (Lovett *et al.*, 1991; Parimoo *et al.*, 1991). A minimum of 10 transcription units were identified, only 5 of which had been previously identified as human genes. All of these units are confined to a 320-kb region of genomic DNA positioned proximal to D17S855 (*BRCA1*) and distal to D17S776.

MATERIALS AND METHODS

Preparation and isolation of genomic DNAs. HSD 1 and HSD2 YAC clones were isolated from pools of CEPH YACs by application of the polymerase chain reaction (PCR) to yield a product of 236 bp with the oligonucleotides ED1 (5′ TTGGCCGTACGTCTGGCTTCA) and ED2 (5′ GTCCAGCTGCAACGTCTCCAG) of *HSD17B1*. Positive clones were verified by hybridization with the cDNAs, hpE$_2$DH216 and hpE$_2$DH36 (Luu-The *et al.*, 1989, 1990). Yeast cultures were propagated, and rapid genomic (Scherer and Tsui, 1991) and total chromosome DNA was isolated as described (Rommens *et al.*, 1994). Separated artificial chromosomes were obtained by electrophoresis with the CHEF-DR II pulsed-field electrophoresis system (Bio-Rad Labs) under standard conditions. They were transferred to Hybond-N nylon membrane (Amersham) and immobilized by UV cross linking for the selection experiments (Rommens *et al.*, 1994).

Genomic DNAs were isolated from human leukocytes or from rodent and rodent hybrid cell lines by established procedures (Miller *et al.*, 1988). The mouse–human hybrid cell line GM10498A was obtained from the NIGMS Human Genetic Mutant Cell Repository, Camden, New Jersey. Restriction digestions were carried out according to the enzyme supplier's recommendations. Electrophoresis, blotting, and hybridization with Hybond-N membrane (Amersham) were carried out with standard procedures (Sambrook *et al.*, 1989). The blots were washed with the stringency achieved with 0.2× SSC with 0.1% SDS at 60°C. The blots were exposed to X-Omat film (Kodak) for 2–72 h at −70°C with one intensifying screen.

For long-range physical mapping, high-molecular-weight DNA embedded in agarose of the cell line GM10498A or of whole yeast chromosomes was digested with rare-cutting restriction enzymes. Products were separated by pulsed-field gel electrophoresis and transferred to nylon membrane for hybridizations as previously described (Church and Gilbert, 1984; Rommens *et al.*, 1989).

cDNA preparation and selection. Randomly primed cDNA was prepared from poly(A)$^+$ RNA of mammary gland, ovary, and placenta tissues and from total RNA of the Caco-2 cell line (ATCC HTB 37), and selection was carried out for two consecutive rounds of hybridization to immobilized artificial chromosome DNA of HSD1, HSD2, and HSD3 as described previously (Rommens *et al.*, 1993, 1994). Hybridizing cDNA was collected, passed over a G50 Fine Sephadex column, and amplified. The products were then digested with *Eco*RI, size selected on agarose gels, and ligated into pBluescript (Stratagene) that had been digested with *Eco*RI and treated with calf alkaline phosphatase (Boehringer Mannheim). Ligation products were transformed into competent DH5α *Escherichia coli* cells (Life Technologies, Inc.).

Characterization of retrieved cDNAs. Individual colonies from each ligation were picked and gridded onto an agar plate containing 100 μg ampicillin for ordering, prescreening, and storage. Grown colonies were transferred onto Biotrans membranes (ICN) and hybridized with radiolabeled cDNA and *HSD17B1* cDNA segments. Plasmids from 20 to 40 clones from each selection experiment that did not hybridize in the prescreenings were isolated for further analysis. Of clones tested, approximately 85% mapped appropriately to the starting YAC clones. The membranes of the grids were also repeatedly hybridized with individual cDNA clones to determine overlapping or redundant clones.

RNA preparation and hybridization analysis. Total RNA was isolated by centrifugation through a CsCl cushion following tissue disruption in a Polytron homogenizer or resuspension of pelleted cells from cultured lines in guanidinium isothiocyanate (MacDonald *et al.*, 1987). Poly(A)$^+$ RNA or blots containing poly(A)$^+$ RNA were purchased and used according to the supplier's recommendation (Clontech Labs). For hybridization analysis, RNA was fractionated on agarose gels (1%) containing 0.6 *M* formaldehyde and transferred to Hybond-N membrane (Amersham). Following cross-linking with UV radiation, blots were repeatedly hybridized with cDNA fragments radiolabeled by random priming (Feinberg and Vogelstein, 1983). Final conditions for washing were as for genomic blots. Blots were exposed to X-Omat AR film for 1–12 days with a single intensifying screen at −70°C. Probes were removed between hybridizations by heating membranes to 95°C in 0.02× SSC and 0.01% SDS.

cDNA library screening. To provide a preliminary screen, PCR was performed using specific oligonucleotides derived from the sequences of each of the retrieved cDNA clones (individual clones were named GT (number)) with DNA from four λgt11 cDNA libraries (Clontech Labs) prepared from human mammary gland (HL 1037b), human breast (HL 1061b), ZR-75-1 human breast cancer cells (HL 1059b), and human ovary (HL 1098b). Subsequently, 13 cDNA fragments were radiolabeled and used to screen 0.5–2.0 × 10^6 pfu as indicated in Table 1. Prehybridization and hybridization were performed as followed at 42°C in 50% formamide, 5× SSPE, 0.1% SDS, 5× Denhardt's mixture, 0.2 mg/ml denatured salmon testis DNA, and 0.002 mg/ml poly(A). Dextran sulfate (4%, v/v) was included in the hybridization solution. Membranes were washed under the same conditions as those for the genomic blots and exposed overnight at −80°C. The positive phages were retested for a second and third screening, if necessary, to obtain purified plaques for subcloning into pBluescript SK (Stratagene) vector. PCR was applied to the positive clones with the universal oligonucleotides of the λgt11 vector and with the oligonucleotides of the GT clones to determine insert size.

DNA sequencing and sequence analysis. The sequences of both strands of the retrieved cDNAs were obtained with the dideoxy chain termination method (Sanger *et al.*, 1977) using the T7 sequencing kit (Pharmacia Biotech Inc.). Initial sequence analysis including assembly, cross alignment, and identification of open reading frames was performed using the Microgenie sequence analysis software (IntelliGenetics, Inc.). The sequences were then subjected to alignment searches through the nonredundant GenBank/EMBL and dbEST databases using GCG's WORDSEARCH and FastA programs from the University of Wisconsin and the BLAST programs from the National Center for Biotechnology Information (NCBI) (Altschul *et al.*, 1990)

using the NCBI e-mail BLAST servers. Those cDNAs that were not identical to previously characterized cDNAs of proteins of known function were tested for the presence of putative exons using GRAIL (Gene Recognition Analysis Internet Link) (UnerBacher and Mural, 1991; Shah *et al.,* 1994). Potential coding regions were examined for known peptide motifs using the BLOCKS database (version 8.0) e-mail searcher at the Fred Hutchinson Cancer Research Center (Henikoff and Henikoff, 1994). Sequences from the GT clones and the corresponding longer cDNA clones were submitted to the dbEST database at NCBI under the GenBank accession numbers listed in Table 1.

RESULTS

Genomic Clones

The region surrounding HSD17B was cloned by screening pools of the CEPH YAC library by PCR using oligonucleotides corresponding to the *HSD17B1* cDNA (Luu-The *et al.,* 1989). Several clones were found to be positive, but they appeared unstable, with significant rearrangement and internal deletions occurring even within single growth cycles. Two YAC clones were confirmed by hybridization with cDNAs, 279F8 (HSD1, 210 kb) and 368C7 (HSD2, 700 kb), and were used to retrieve cDNAs and to build the physical map. It was determined that HSD2 was chimeric and that a large portion of the end containing the right YAC arm was found to originate from chromosome 16 (data not shown). An additional YAC clone, known to contain the HSD17B locus, 26F3 (HSD3) of 290 kb was also used for analysis and mapping. The overlap and alignment of these YACs are shown in Fig. 1.

Retrieval of cDNA Fragments and Characterization by Physical Mapping

A pool of primary cDNAs prepared from poly(A)$^+$ RNA of human mammary gland, ovary, and placenta as well as from total RNA of the Caco-2 cell line was hybridized to purified and immobilized YAC clones (Rommens *et al.,* 1993, 1994).

Initial analysis of the cDNA clones involved a prescreen for ribosomal sequences. Approximately 10% of the clones were eliminated as they hybridized strongly with radiolabeled cDNA obtained from total RNA. Fewer than 3% of the remaining clones were found to be derived from the HSD17B locus as identified by hybridization to cDNA fragments of *HSD17B1* (Luu-The *et al.,* 1989). One positive clone, GT207, was analyzed by sequencing for further verification.

The retrieved cDNA fragments that were not eliminated in the prescreenings were then verified to originate from the starting YAC by hybridization to restriction digests of DNA of the yeast clones; six examples are shown in Fig. 2A. The clones were tentatively assigned into groups based on the overlapping and nonoverlapping intervals of the YAC clones. The position of each clone was subsequently verified by long-range physical mapping both with the artificial chromosomes and with genomic DNA (data not shown).

That each of the cDNAs also originated from chromosome 17 was confirmed by hybridization to a mouse hybrid cell line that contains chromosome 17 as its only human material; six examples are shown in Fig. 2B. With one exception, the *Eco*RI restriction fragments detected in the yeast DNA corresponded to those detected in the chromosome 17 hybrid and in human DNA. The cDNA clone, GT209, displayed an interesting pattern that is shown in Fig. 2. Subsequent investigation indicated that this clone mapped to the left arm boundary and was present only partially in HSD1. Further, the smallest *Eco*RI band that was detected in HSD2 and HSD3 is polymorphic (data not shown). Two groups of clones, (1) GT208 (shown in Fig. 2B), GT214, GT217, and GT252 and (2) GT218, GT298, GT304, and GT305, and the singlet clone GT247 showed strong hybridization to chromosome 17 but also yielded additional bands with comparable or weaker intensity in total human DNA.

RNA Hybridization

Retrieved cDNA clones were hybridized to total RNA of a series of different tissues and cell lines; examples are shown in Fig. 3. The results from the RNA hybridization together with the initial physical mapping permitted groups of cDNAs to be assigned to individual transcription units. For example, GT208 and GT217 did not overlap in sequence but originated from a common interval of the HSD2 YAC that did not include portions of HSD1 or HSD3 as shown in Fig. 1. Both hybridized to a common mRNA species of 3.2 kb in several tissues, as shown in Fig. 3. Subsequently, a longer cDNA clone identified from a mammary gland library confirmed the unit overlap as it contained both of the smaller cDNA fragments.

Comparable analyses of each of the cDNA fragments yielded evidence for at least 10 transcription units within 320 kb of genomic DNA. The groups of clones and their relative map positions of the individual units have been summarized in Fig. 1 and are shown numbered 1 through 10. The precise order for transcripts 4 and 5 or for 7 and 8 could not be determined from the long-range mapping information obtained. Transcription unit 9 corresponded to *HSD17B1* (Luu-The *et al.,* 1989). RNA hybridization for transcription units 1, 2, 3, 6, 7, 8, and 10 is shown in Fig. 3. mRNAs corresponding to GT219, GT227, GT232, and GT235 were not observed by RNA hybridization analysis, so no corresponding transcription units were assigned. Three of these four clones map to the same physical interval as shown in Fig. 1.

One group of clones, including GT218 and GT305, hybridized to at least two discrete messages of 4.0 and 1.9 kb. Further, the clone GT305 also detected a mRNA of 1.0 kb specifically in the brain; see Fig. 3. These

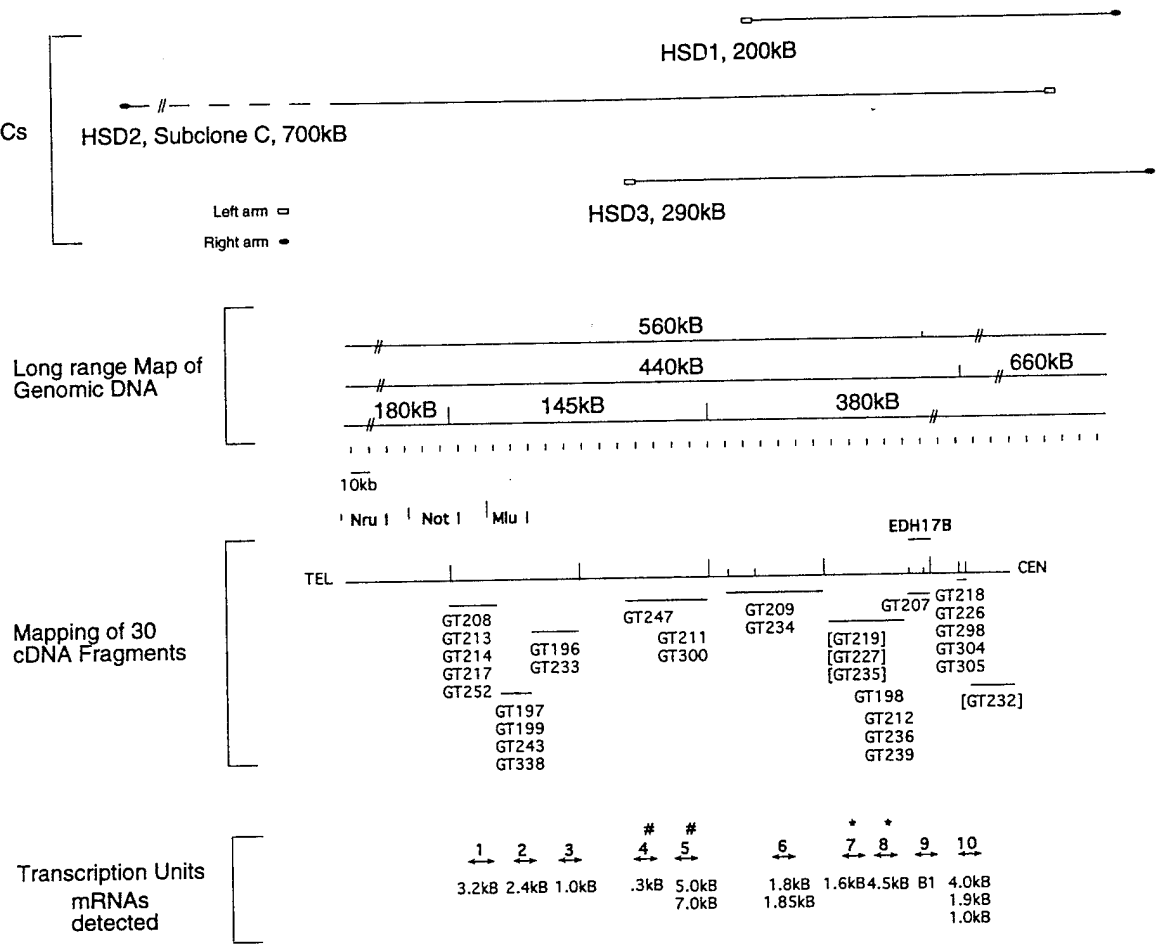


FIG. 1. Physical and transcript map at the HSD17B locus. The overlapping YAC clones containing the HSD17B locus and the detailed physical map that was generated are shown aligned to a long-range physical map of human DNA. The positions of the restriction sites for *Mlu*I, *Not*I, and *Nru*I were determined by single and double digestion combinations. Because of absence of methylation, additional restriction sites were noted in the YACs. The map interval of each cDNA clone is indicated by the horizontal lines shown drawn above the listed clones. These intervals were determined by overlap of the three YACs or by the analysis of additional subclones of the HSD2 YAC that contained deletions; only the longest subclone of HSD2 is shown. Map position and RNA hybridization results were used to obtain the transcript map shown below the groups of clones. Pound signs and asterisks indicate that transcript order within the single physical intervals could not be determined. The clones GT219, GT227, GT232, and GT235 shown enclosed by brackets were not assigned to transcription units.

mRNAs were grouped as transcription unit 10. Although these clones hybridized very strongly to single *Eco*RI and *Not*I restriction fragments in YAC genomic DNA, a more complex hybridization pattern was observed with human DNA. Hybridization to a hybrid mapping panel indicated that at least one copy of these selected sequences did originate from chromosome 17 but that at least four additional copies were present elsewhere in the genome, specifically on chromosomes 1, 6, 11, and 12 (data not shown).

Two selected cDNAs, GT211 and GT198, that correspond to transcription units 5 and 7 were also assessed by hybridization to poly(A)$^+$ RNA, as shown in Fig. 4, to obtain a more complete analysis of pattern of expression. They are described in subsequent sections. The two mRNAs of 5.0 and 6.5 kb of transcription unit 5 (Fig. 4A) that are detected with GT211 may be due to alternative splicing or to differential polyadenylation but are likely to be derived from a single transcriptional unit. The pattern of hybridization to DNA with GT211 was complex (see Fig. 2B) but was consistent with a single location on chromosome 17. The complexity resulted from the strong cross-hybridizing bands of rodent DNA and because at least five exons were present in the GT211 cDNA fragment. Each of the human bands was present on chromosome 17 and on both HSD1 and HSD2 YACs, but was absent on HSD3; see Fig. 2A.

Longer cDNA clones were obtained for selected cDNA fragments as summarized in Table 1. Presence in the cDNA libraries was initially established by amplification with specific oligonucleotides designed from the sequences of the selected fragments.

cDNAs of Three Known Human Genes

The sequences of selected cDNAs and of isolated corresponding cDNA clones were determined and ana-

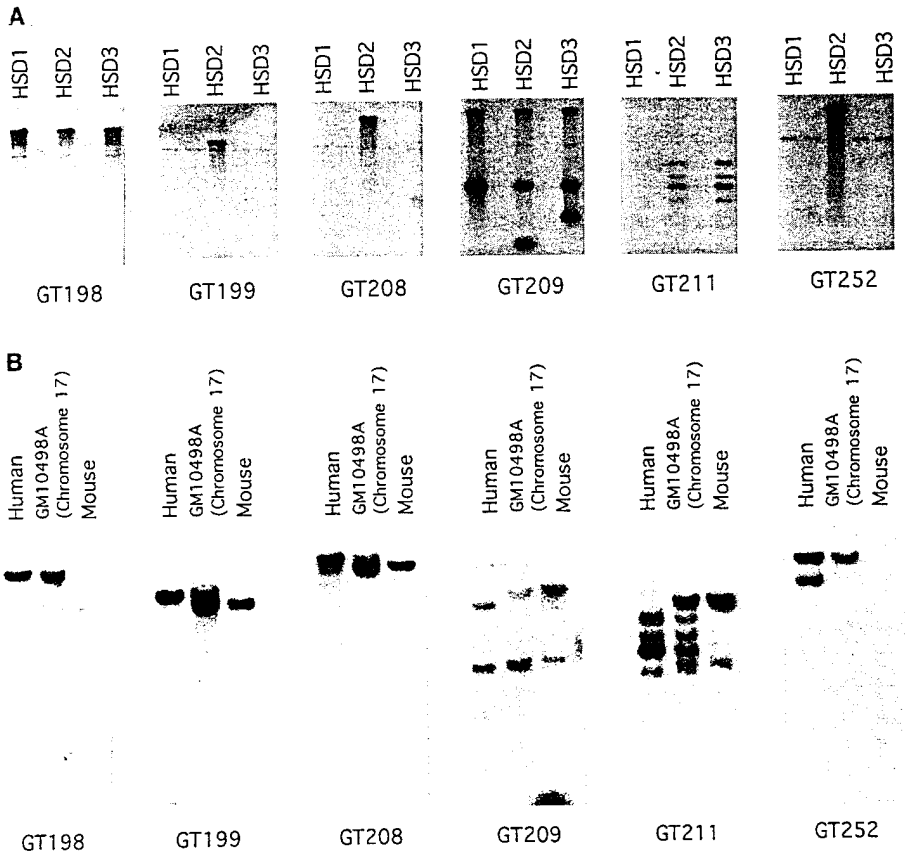


FIG. 2. Verification of origin of selected cDNAs to genomic DNA. Six cDNAs are shown hybridized to (**A**) 0.5 μg of total genomic yeast DNA of YAC clones HSD1, HSD2, and HSD3 and (**B**) 10 μg of mammalian DNA that had been digested with *Eco*RI restriction enzyme. The DNAs were separated on 0.7% agarose gels and transferred to nylon membrane. The lanes between the labeled yeast DNAs are control samples that correspond to YAC clones that do not overlap with the HSD17B locus. The mammalian DNA was prepared from human leukocyte, mouse–human hybrid (GM10498A), and mouse L cell line. The hybrid mouse cell line contains chromosome 17 as its only human material. cDNAs (e.g., GT252) could not be assigned definitively to chromosome 17 when hybridizing bands appeared in human DNA that did not appear in the hybrid cell line. Cross-species hybridization is noted with several of the retrieved cDNAs.

lyzed for coding potential and compared to sequences of the public databases. Three of the 10 transcription units could be readily identified; they corresponded to the genes for the Ki nuclear autoantigen (Nikaido *et al.,* 1990; Albertson *et al.,* 1994b), γ-tubulin (Zheng *et al.,* 1991), and *HSD17B1* (Luu-The *et al.,* 1989). One hundred percent identity of sequence was observed between the retrieved cDNA fragments and the human gene sequences available. The details are given in Table 1. These genes are shown as transcription units 1, 6, and 9, respectively, in Fig. 1.

One interesting feature was noted for the mRNA that corresponded to the γ-tubulin gene. The 1.8-kb transcript, detected by GT234 in Fig. 3, was found in all RNA samples tested but appeared as a doublet band in the brain. This suggests the existence of a tissue-specific alternative splicing or differential polyadenylation.

Identification of Human Homologues of Two Rat Genes

Tentative identification of two other transcription units, 4 and 10, could be achieved by alignment of translated cDNA sequences to rat genes. The nucleotide sequence of the predicted coding region of GT247, transcription unit 4, shares 90% identity with that of the rat ribosomal gene, *L34.* Significant similarity was also observed between the 117-amino-acid putative peptide encoded and the 117-amino-acid peptide of *L34* (Aoyama *et al.,* 1989). With the exception of the 13 consecutive residues from amino acid position 75 through and including position 88, due to a deletion of a T at codon 74 and an insertion of a T at codon 88, only 5 of the remaining 104 amino acids were not identical. Further, as shown in Fig. 5, the positions of the initiation and termination codons were identical.

The sequence corresponding to transcription unit 10, which was compiled from overlapping sequences of GT218, GT226, GT298, GT304, and GT305 and several cDNA clones isolated from a mammary gland cDNA library, shared 72% sequence identity with the coding region of a rat liver tyrosine phosphatase cDNA (Diamond *et al.,* 1994). Translation of the composite sequence yielded an apparent incomplete open reading frame of 154 amino acids. Ninety percent were identical

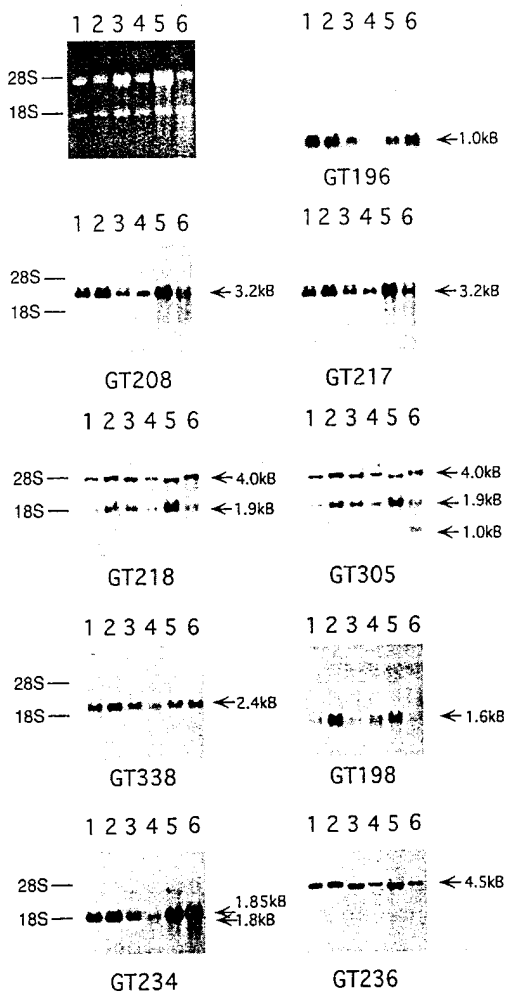


FIG. 3. Hybridization of cDNAs to total RNA. Nine retrieved cDNAs are shown hybridized to 10 μg of total RNA of (**1**) Caco-2 (intestinal cell line), (**2**) BT-474 (breast carcinoma cell line), (**3**) fibroblast, (**4**) HL60, (**5**) lymphoblast, and (**6**) frontal cortex of adult brain as indicated. They demonstrate 7 of the 10 transcription units shown in Fig. 1. The integrity of the RNA is shown in the gel stained with ethidium bromide (upper left). The positions of the ribosomal RNAs, 18S and 28S, are indicated to the left. Sizes of the transcripts in kilobases (kb) are given to the right of each panel.

to the corresponding portion of the 173 amino acids of the nuclear tyrosine phosphatase of the immediate-early *PRL-1* gene of rat (Diamond *et al.,* 1994). These results indicate that the cDNAs corresponding to transcription units 4 and 10 encode the human homologues for the rat ribosomal *L34* and *PRL-1* genes, respectively.

Identification of a Putative Transcription Factor

The sequence of 818 nucleotides of clone GT211, which identified transcription unit 5, did not align to human or mammalian sequences. Alignment of the translated amino acid sequence did reveal segments of striking amino acid similarity with translated portions of the *Enhancer of zeste* gene, *E(z),* of *Drosophila melanogaster* (Jones and Gelbart, 1993). In two stretches corresponding to the first 42 and the last 50 amino acids in the alignment shown in Fig. 6, 57 and 70% identities were observed. Also, the latter region of identity included a putative nuclear localization signal as indicated, followed by the consistent spacing and conservation of 5 of 5 cysteines within the following 41 amino acids. Portions of this human homologue have also been reported in a preliminary map generated by others (Friedman *et al.,* 1994b).

E(z) in *D. melanogaster* is involved in the negative regulation of the segment identity genes of the *Antennapedia* (ANT-C) and *Bithorax* (BX-C) gene complexes. This gene exhibits two mRNA species, 2.4 and 3.3 kb, that are regulated in a development-specific manner. The smaller mRNA is expressed during embryo, pupae, and adult stages, while the larger 3.3-kb mRNA is not expressed in the embryo, but is expressed in the pupae and adult stages with higher abundance in males (Jones and Gelbart, 1993). These transcripts are notably smaller than the two mRNAs of 5 and 6.5 kb detected with GT211 in many adult tissues (Fig. 4A). There appears to be some tissue-specific regulation of the individual human messages; for example, skeletal muscle displayed the smaller mRNA most prominently, while nearly equal levels were detected in spleen and thymus. Expression of this gene was also observed in fetal tissues including brain, lung, liver, and kidney (data not shown).

Novel Genes and Identities with Expressed Sequence Tags

Identification of the roles of the genes corresponding to the four remaining transcription units was limited, as distinctive peptide motifs were not recognized. Transcription unit 3, detected by GT196 and GT233, corresponded to a small mRNA of 1.0 kb. Neither the nucleotide nor the translated amino acid sequences aligned in the public databases, despite the notably abundant expression levels observed in human tissues including adult brain (Fig. 3). Further characterization was confounded by observing that all nine cDNA clones isolated from the mammary gland cDNA library were chimeric.

Segments of the retrieved cDNAs of transcription units 2, 7, and 8 aligned with near identity to expressed sequence tags determined from cDNA libraries of the HL60 cell line or of a 72-day-old human female brain, as detailed in Table 1. Discrete mRNAs of 2.4 and 4.5 kb were detected for transcription units 2 and 8, respectively. Both appear to be relatively abundant in all tissues examined; see the hybridization results with GT338 and GT236 in Fig. 3. The full-length cDNA corresponding to transcription unit 7 is discussed below.

Isolation of Full-Length cDNA Clones for Transcription Unit 7

GT198 of transcription unit 7 revealed a message of 1.6 kb with dramatic differences in expression levels between

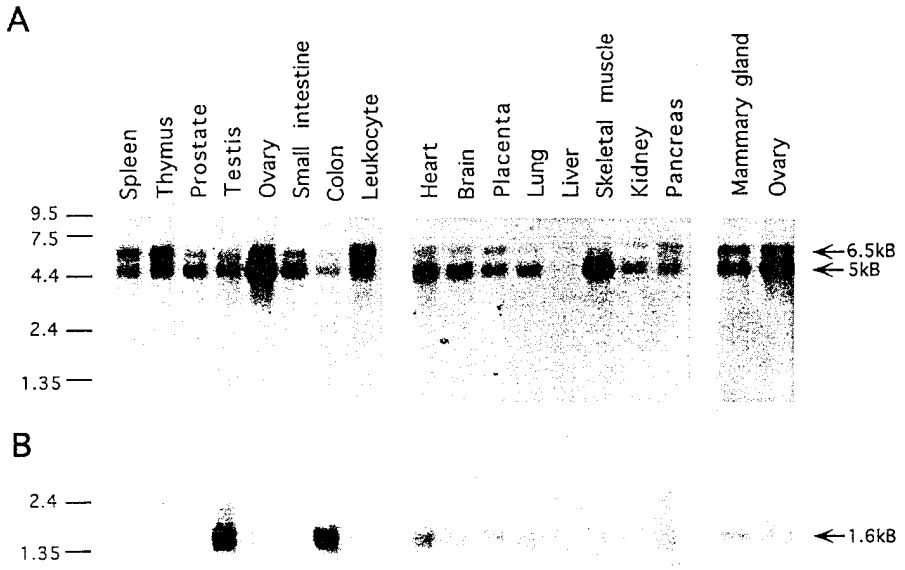


FIG. 4. Hybridization of cDNAs to poly(A)$^+$ RNA. GT211 (**A**) and GT198 (**B**) are shown hybridized to 2 μg of poly(A)$^+$ RNAs of the tissues indicated. The blots were obtained from Clonetech Labs. The sizes of the markers listed to the left and the sizes of mRNAs, as indicated to the right, are given in kilobases (kb). In **B,** the exposure time of the blots of the first series of tissues including spleen, thymus, prostate, testis, ovary, small intestine, colon, and leukocyte was 0.7 days, while the exposure time for the remaining lanes shown was 8 days.

tissues. Marked abundance was observed in three samples including the BT-474 cell line (Fig. 3), and testis and colon tissues (Fig. 4B). This contrasted with the barely detectable level observed in a series of cell lines including one derived from a colon carcinoma (Caco-2) and several other tissues including normal mammary gland. As demonstrated in the central and right portions of Fig. 4B, low-level expression was observed upon prolonged exposure of the blots in all tissues tested. Low-level expression was also observed in fetal samples of liver, kidney, lung, and brain (data not shown).

Given the interesting expression pattern, the full-length cDNA for GT198 was obtained by repeated screenings of human breast cancer cell (ZR-75-1) cDNA library (Fig. 7). The length of this cDNA was consistent with the size of the mRNA observed. From the nucleotide sequence shown in Fig. 7B, the longest putative open reading frame included 203 amino acids. This open reading frame orientation was consistent across splicing boundaries of the introns found in the initial GT198 cDNA fragment and is preceded by a 5′ untranslated region of 133 nucleotides. Further, a polyadenylation consensus signal was located 17 nucleotides upstream from the poly(A) tail. The expressed sequence tags that could be aligned (see Table 1) corresponded to 287 nucleotides of the 3′-untranslated region (EST Z38575) and to the 99 nucleotides of codons 73 through 106 (EST 42353) within the open reading frame. It was, however, not possible to gain insight into the function of this gene as no known protein motifs were recognized.

DISCUSSION

All of the transcripts have not been explored in detail, and transcription units were defined only when evidence for RNA hybridization was detected. Four of the selected cDNA fragments did not detect discrete mRNAs in total RNA and thus were not aligned into transcription units. Based on analysis by GRAIL, they did not reveal strong potential for coding sequence, as complete open reading frames could not be detected in these clones. However, evidence that they corresponded to additional genes was observed, as three of these four fragments displayed cross-species hybridization. Further, as indicated in Table 1, GT232 aligned to an expressed sequence tag. The sensitivity of direct selection is very high, and it is possible that these fragments correspond to genes that are restricted in their expression pattern or are expressed at very low level, below the detection limit of RNA hybridization. As the clones GT219, GT227, and GT235 mapped to the same interval, it is possible that they are distinct portions of the same rare transcript.

Only four tissues were used as the cDNA sources of the selection experiments, so it cannot be determined whether all transcribed sequences from the region analyzed have been identified. However, the region examined was found to be relatively rich in genes. These included genes that are expressed at relatively low levels as well as those that appear broadly and abundantly expressed in several tissues. A summation of the sizes of all of the transcripts detected suggests that up to 25–30 kb of the genomic DNA from the 320-kb region analyzed is present in mature transcripts. It was notable that long range mapping with rare cutting enzymes did not predict the high density of genes found.

While initial findings based on the tissue-specific expression pattern indicated that GT198 may have been

TABLE 1

Summary of cDNA Clones

Retrieved GT clones			Isolated breast cDNA subclones (HL1061b cDNA library)				Transcription unit		
Name	Size (bp)	cDNA source	Isolated cDNA Clones	Size of isolated clones (kb)	Size of longest contig (kb)	GenBank accession number	Number	Transcript size (kb)	Homologies
GT208	554	Caco-2	4	1.0, 1.7, 2.0 and 3.8	0.6	—	1	3.2	Both contigs identical to human Ki antigen cDNA (GenBank Accession No. U11292) GT208 contig ≡ U11292 bases 735–1289 GT217 contig ≡ U11292 bases 1860–2442 EST M85569 ≡ GT217 contig bases 560–355 GT217 homolgous to part of mouse EST #L26609
GT213	554	Caco-2							
GT214	579	Caco-2	2	0.7 and 3.8	0.6	—			
GT217	272	Placenta							
GT252		Caco-2							
GT197	628	Caco-2	3	0.8, 1.5, and 2.9	1.5	L38932	2	2.4	EST D20904 ≡ GT197 contig bases 1377–1530
GT199	860	Caco-2	3	0.85, 1.6, and 2.8					
GT243	627	Caco-2							
GT338	657	Placenta							
GT196	486	Caco-2	9	0.7, 0.750, 1.0, and 1.8	0.6	L38939	3	1.0	No homology
GT233	625	Caco-2							
GT247	392	Ovary	—	—	—	L38941	4	0.3	Human homolog of rat ribosomal protein L34 cDNA (GenBank Accession No. X14401)
GT211	819	Caco-2	5	1.0, 1.6, 1.9, 2.2, and 3.4	0.8	L38934	5	5.0 and 6.5	Similarity with part of *D. melanogaster Enhancer of zeste (E(z))* cDNA (GenBank Accession No. U00180)
GT300	607	Ovary							
GT209	631	Caco-2	—	—	—	—	6	1.8	The two original clones identical to human γ-tubulin cDNA (GenBank Accession No. M61764) GT209 bases 8–607 ≡ M61764 bases 33–631 GT234 ≡ M61764 bases 716–1559
GT234	844	Caco-2							
GT198	418	Caco-2	2[a]	0.9 and 1.5	1.5	L38933	7	1.6	EST Z42353 bases 103–201 ≡ GT198 contig bases 353–451 EST Z38575 ≡ GT198 contig bases 1439–1153
GT212	374	Caco-2	4	0.8, 1.2, 2.1, and 4.2	1.2	L38935	8	4.5	EST T17366 ≡ GT212 contig bases 1–322
GT236	990	Caco-2							
GT339	374	Caco-2							
GT207	1216	Caco-2	—	—	—	—	9	2.3 and 1.3	Identical to human 17β-HSD type I cDNA (GenBank Accession No. X13440 bases 35–1250)
GT218	615	Caco-2	1	2.9	1.1	L39000	10	1.9 and 4.0	GT218 contig is the human homolog of a rat tyrosine phosphatase cDNA (GenBank Accession No. L27843) EST Z30155 ≡ GT218 contig bases 583–882 EST T10429 ≡ GT218 contig bases 399–609 GT305 contig bases 1–136 ≡ Z15428 bases 200–336
GT226	~300	Caco-2	6	1.2, 1.4, 1.6, 2.1, 2.3, and 2.4					
GT298	~750	Caco-2							
GT304	~700	Caco-2							
GT305	~418	Mammary gland							
GT219	535	Caco-2	—	—	—	L38936	—	n.d.	No homology
GT227	303	Ovary	1	1.3	0.3	L38937	—	n.d.	No homology
GT232	423	Ovary	1	2.9	0.5	L38938	—	n.d.	EST Z40888 ≡ GT232 contig bases 472–180
GT235	537	Caco-2	10	0.8, 1.2, 1.3, 1.4, 2.0, and 2.3	0.7	L38940	—	n.d.	No homology

Note. A summary of the characterization of the retrieved clones and the corresponding isolated cDNAs from cDNA libraries are listed. The clones are grouped corresponding to the transcription units shown in Fig. 1. For the selected clones, the tissue source could be identified by the sequence tag that had been incorporated during the synthesis of cDNA. The longest length of the cDNA "contig" given in the sixth column was obtained by compiling overlapping sequences of selected clones and longer cDNAs. The entire contig sequence was assigned the GenBank accession numbers that are indicated. n.d. denotes none detected.

[a] The cDNA library screened for GT198 was made from the ZR-75-1 human breast carcinoma cell line (Clonetech Labs, HL1059b).

an interesting candidate gene for involvement in breast cancer, the recent identification of disease-associated mutations in a putative transcription factor (Miki *et al.*, 1994; Futreal *et al.*, 1994a; Castilla *et al.*, 1994; Simard *et al.*, 1994; Friedman *et al.*, 1994a) argue against this being BRCA1. The region surrounding GT198 is not amplified in the BT-474 cell line (data not shown), but the interesting expression pattern ob-

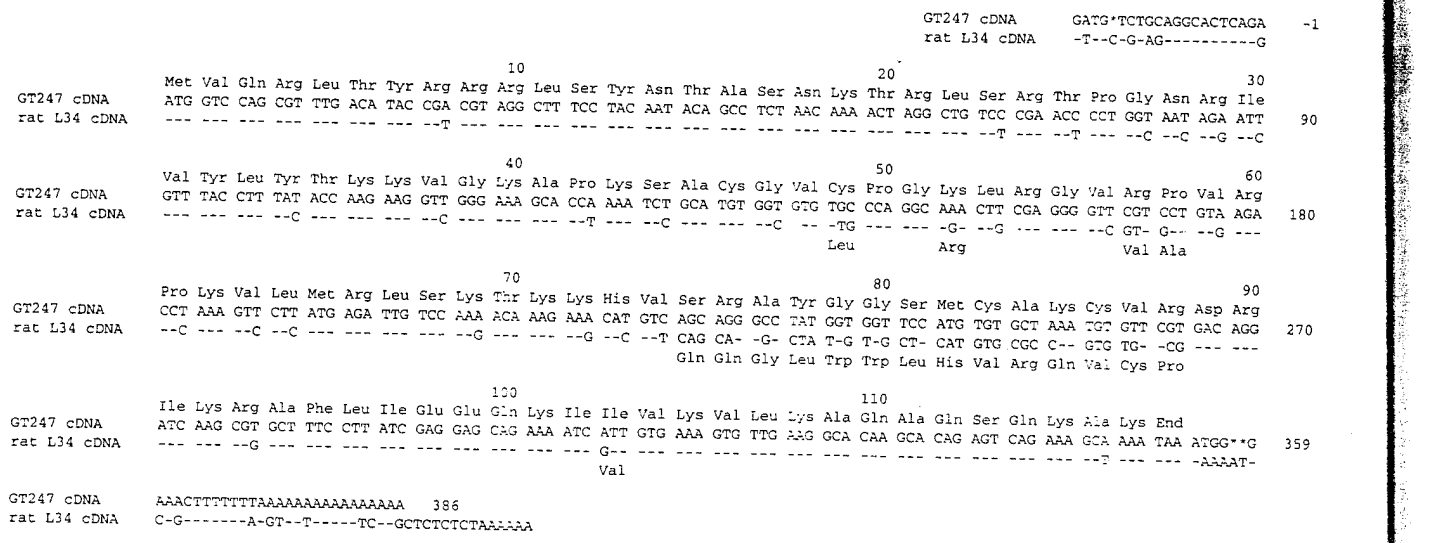


FIG. 5. Comparison of nucleotide and predicted amino acid sequences of GT247 and rat L34. Comparison of the complete nucleotide and translated sequence of GT247 and the rat L34 ribosomal protein is shown. Nucleotide numbering is indicated to the right of each line of GT247 sequence. Amino acids, indicated above, are shown numbered according to the putative ATG initiation codon of the rat L34 ribosomal protein. Differences in sequence and amino acids are noted, while identical nucleotides and amino acids are indicated by dashes and blanks, respectively.

served suggests that additional analysis of this gene may be warranted in sporadic breast cancers.

The complete characterization of the putative transcription factor corresponding to the transcription unit detected with GT211 will also be of interest. While a role for this gene cannot yet be accurately postulated in human, the role of *E(z)* in *Drosophila* is to provide a negative controlling component or feature of transcription of the *Antennapedia* and *Bithorax* gene complexes. This is thought to involve the counteraction of the *E(z)* gene and the trithorax gene (*trx*) products. It is interesting to note that the human homologue of the *trx* gene, which maps to chromosome 11q23, has been shown to be involved in translocations in acute leukemias (Djabali *et al.*, 1992; Tkachuk *et al.*, 1992). We speculate that it is possible that abnormal depletion or inactivation of a negative component of transcription is comparable to the abnormal activation of a positive component of transcription that occurs via translocation. Consequently, aberrations of activities of the transcription unit detected by GT211 may also be involved in some cancers.

Two sets of clones appeared to have features that emphasized the limitations of direct selection procedures for the identification of genes. For example, the sequencing analysis of GT198 indicated that this selected clone likely contained at least a portion of one intron and another intact intron. This was determined by alignment of sequence to longer cDNAs that were subsequently isolated from a ZR-75-1 human breast cancer cell cDNA library. The alignment of identical sequence ended abruptly within the cDNA, and se-

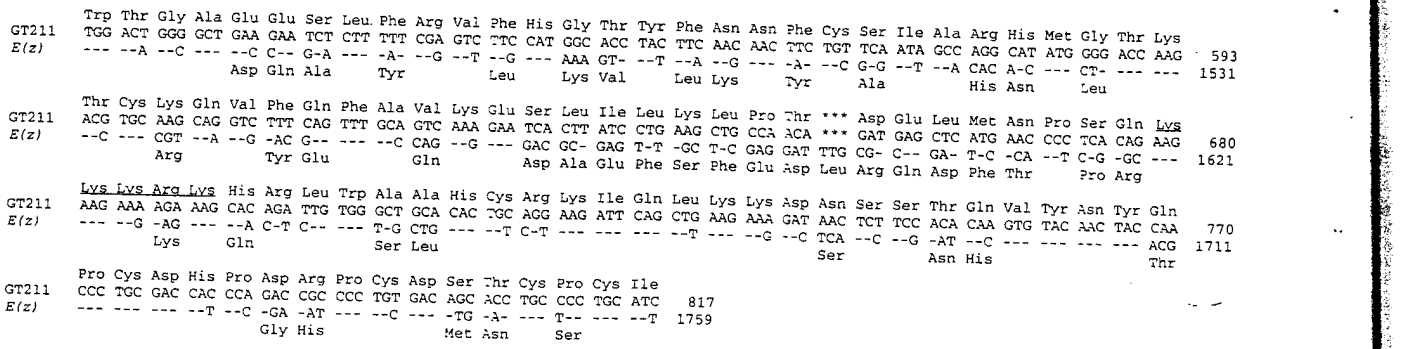
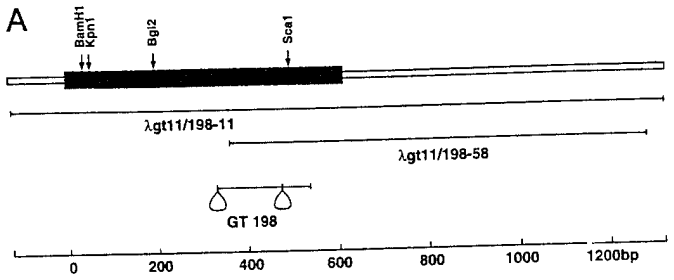


FIG. 6. Alignment of GT211 with *Enhancer of zeste*. Alignment of the nucleotide and predicted amino acid sequence of GT211 with a portion of the *E(z)* gene of *D. melanogaster* is shown. Nucleotides are numbered to the right of each line of sequences according to the GenBank entries of GT211 (L38934) and of *E(z)* (U00180). The sequence of GT211 cDNA is given; identical and different *E(z)* nucleotides are indicated by dashes or the corresponding nucleotides, respectively. Only the different amino acid residues of *E(z)* are indicated. Asterisks indicate a gap in the human sequence that allows maximum alignment of identical or similar amino acids. A putative nuclear localization signal is underlined.



B

```
                                                                                                                ATCAAGGTGATCCC   -120
AAAACGAACCAACAGACCAGGCATCAGCACAACAGACCGGGGTTTTCCACGAGCCCGCTACCGCGCCCCGGACCACCAACTACAACGTCCGGCTTTCTGAGTTGGGTGGCGGGAAAGGCG     -1

ATG AGT AAA GGC CGG GCA GAA GCT GCG GGA GCC GCC GGG ATC CTC CTG AGG TAC CTG CAG GAG CAG AAC CGG CCC TAC AGC TCC CAG GAT      90
Met Ser Lys Gly Arg Ala Glu Ala Ala Gly Ala Ala Gly Ile Leu Leu Arg Tyr Leu Gln Glu Gln Asn Arg Pro Tyr Ser Ser Gln Asp      30

GTG TTC GGG AAC CTA CAG CGG GAA CAC GGA CTG GGC AAG GCG GTG GTG GTG AAG ACG CTG GAG CAG CTG GCG CAA CAA GGC AAG ATC AAA     180
Val Phe Gly Asn Leu Gln Arg Glu His Gly Leu Gly Lys Ala Val Val Val Lys Thr Leu Glu Gln Leu Ala Gln Gln Gly Lys Ile Lys      60

GAG AAG ATG TAC GGC AAG CAG AAG ATC TAT TTT GCG GAT CAG GAC CAG TTT GAC ATG GTG AGT GAT GCT GAC CTT CAA GTC CTA GAT GGC     270
Glu Lys Met Tyr Gly Lys Gln Lys Ile Tyr Phe Ala Asp Gln Asp Gln Phe Asp Met Val Ser Asp Ala Asp Leu Gln Val Leu Asp Gly      90

AAA ATC GTG GCC CTC ACT GCT AAG GTG CAG AGC TTG CAG CAG ACG TGC CGC TAC ATG GAG GCT GAG CTC AAG GAA TTA TCT AGT GCC CTG     360
Lys Ile Val Ala Leu Thr Ala Lys Val Gln Ser Leu Gln Gln Thr Cys Arg Tyr Met Glu Ala Glu Leu Lys Glu Leu Ser Ser Ala Leu     120

ACC ACA CCA GAG ATG CAG AAA GAA ATC CAG GAG TTA AAG AAG GAA TGC GCT GGC TAC AGA GAG AGA TTG AAG AAC ATT AAA GCA GCT ACC     450
Thr Thr Pro Glu Met Gln Lys Glu Ile Gln Glu Leu Lys Lys Glu Cys Ala Gly Tyr Arg Glu Arg Leu Lys Asn Ile Lys Ala Ala Thr     150

AAT CAT GTG ACT CCA GAA GAG AAA GAG CAG GTG TAC AGA GAG AGG CAG AAG TAC TGT AAG GAG TGG AGG GAA GAG GAA GAG GAT GGC TAC     540
Asn His Val Thr Pro Glu Glu Lys Glu Gln Val Tyr Arg Glu Arg Gln Lys Tyr Cys Lys Glu Trp Arg Glu Glu Glu Glu Asp Gly Tyr     180

AGA GCT GTC TTG ATG CAA TAC TTG AAG GAT ACC CCA AGA GCA AGA AGC AGT TCT TTG AGG AAG TTG GGA TAG AGACGGATGAAGATTACAACGTC     635
Arg Ala Val Leu Met Gln Tyr Leu Lys Asp Thr Pro Arg Ala Arg Ser Ser Ser Leu Arg Lys Leu Gly End

ACACTCCCAGACCCCTGAGGGGCCCACGGTCAGGACTGGTGGGGACTGCAGGATGTCAGAAGAGTGAGATGTCTTGCACTGGCTACCTTGTTTTTGGTGGCTTTTGTTGTTGTTCCTCT     754
TACTTTTCACTTTAGCAGAGCAGTCAGGAGACAAGCATAAACCAGAGCACTGGGTAGAGAGGATGAGGGCTGGTGGCTGGGGGTAGACCCCACGCATTTCATTGTCTAAATTGCAGTAG     873
CTTGAGGTTAACATTTAGACTTGGAACAATGCTAAAGGAAAGCATTTGGCAATATTTATTATAATTTAATTTTATATAAAAATATTTAATTTCCTCTGGATAGTCAAACCTGCCAGATA     992
TCAAACCTGAGGAAGGCAGAAGTGAATTTGGAGAACTAGGGTAGAGAGAGGTTGCTATAAAACGAGCATTTGGAGGGCCCACGGCTTCACTCAGGACCTGCTGGGCTTGTGTACCCCAG    1111
GAGCCCTTTTAAGTATCTTTTGTACGCTTTTCACCCCACCCCCAAGTCCTGGGAGAAATGCAGGCAACACTGAGACATGGGAGAGGCCAAGATATGCTTGACAGAAAGGGTGATTTTGA    1230
GGCTCAGTTAATATTTCAAAATTGTAACCGTAGCAAAACTGCATTGGTATTTAGAAAAATAAAAAATTTCCAATAAAAAAAAAAAAA   1317
```

FIG. 7. Full-length cDNA for GT198. (**A**) Schematic of the structure of the GT198 and derived cDNA clones. The protein coding region is represented as a black box; open boxes denote 5′- and 3′-noncoding regions. Loops in the original GT198 clone denote intronic sequences. The scale in basepairs is shown below. (**B**) Nucleotide and predicted amino acid sequences of GT198 cDNA. Nucleotide and amino acids are numbered to the right of each line of sequence. The +1 nucleotide corresponds to the first nucleotide of the putative ATG initiation codon. An upstream and in-frame stop codon (TGA) is underlined with a dashed line, and a consensus polyadenylation signal at the 3′ end is underlined with a solid line.

quences consistent with two 3′ acceptor sites and a single 5′ donor site were found in the selected fragment with appropriate orientation. Furthermore, the sequence of the longer cDNA was consistent with the maintenance of open reading frame. We have observed that clones containing introns, which we assume must originate from a rare RNA population, may be favored for selection during the hybridization steps. Our findings have been supported by observing that unprocessed cDNAs are more commonly retrieved when cDNA of total cell RNA is used compared to cDNA of poly(A)$^+$ RNA (Rommens *et al.*, 1994, data not shown). It is unlikely that the introns originate from genomic DNA, as the scheme used to generate cDNA does not involve the ligation of oligonucleotide linkers (Rommens *et al.*, 1993).

A second limitation of direct selection, the consequences of gene families or pseudogenes, was observed by analyzing the transcribed unit defined by clones GT218, GT298, GT304, and GT305. Evidence that these clones detect a gene family was obtained, so it could not be ascertained which of two (detected by GT218) or three (detected by GT305) mRNAs (Fig. 3) detected actually originated from chromosome 17. The hybridization pattern to genomic DNA for this group of clones was complex compared to the YAC DNA or chromosome 17 hybrid cell line. With GT218, three discrete *Eco*RI fragments, suggesting three copies or three exons, were found to hybridize from chromosome 1, but only single fragments hybridized from other chromosomes, including 6, 11, 12, and 17 (data not shown). It is possible that the sequences on chromosome 17 actually correspond to a pseudogene or represent a single member of a highly conserved family, but the issue cannot be resolved from our current data. This would require complete sequencing of the hybridizing genomic regions.

We have compared the transcription map generated from the region surrounding the HSD17B locus to recently described physical (Couch *et al.*, 1995) and transcription maps. These studies used exon trapping, conserved sequence screening, CpG island rescue (Brody *et al.*, 1995), and direct library screening with YAC or cosmid DNA (Brody *et al.*, 1995; Friedman *et al.*, 1995),

in addition to direct cDNA selection methods (Brody *et al.,* 1995; Osborne-Lawrence *et al.,* 1995), as strategies to identify transcribed sequences. Transcription units 1 and 9 (Albertsen *et al.,* 1994b), which had been described earlier, and units 5 and 6 were common to all maps and could be readily identified by alignment of sequences.

Transcription units 2, 3, 7, and 8 could also be assigned to transcripts of comparable sizes in at least two studies by partial overlap with expressed sequences that have been recently entered into public databases (Brody *et al.,* 1995; Friedman *et al.,* 1995). Overall there was agreement on the relative position of each of these transcription units, and supporting information that transcription units 7 and 8 are in the correct order, as shown in Fig. 1, was evident (Couch *et al.,* 1995; Brody *et al.,* 1995).

Of the expressed sequences that were obtained that were not assigned into transcription units, only GT235 (Accession No. L38940) was found to overlap with an exon-trapped clone (Accession No. T27158; Brody *et al.,* 1995) of one study. Unassigned cDNA fragments were isolated in all of the studies, and it is possible that some correspond to other common transcripts that could not be recognized, as different nonoverlapping fragments have been isolated. The most noteworthy difference between the transcription map that we have generated and those of others is the absence of the pseudogene for the high-mobility group protein, *HMG17* (Brody *et al.,* 1995; Friedman *et al.,* 1995; Osborne-Lawrence *et al.,* 1995). We also do not have evidence for either the T-eta transcript (5.0-kb mRNA, Brody *et al.,* 1995) or the 6.5-kb transcript (Friedman *et al.,* 1995) that were obtained but that would map between units 5 and 6, or for a 2.5-kb transcript (Friedman *et al.,* 1995) that would be positioned near transcription unit 7. Alternatively, neither sequence overlap nor mRNA size permitted recognition of either transcription units 4 or 10 in the alternate maps, although it is possible that portions of unit 10 were obtained based on physical position of unassigned exon-trapped clones (Brody *et al.,* 1995). As the different transcripts have not been examined in sufficient detail, it is not possible to know whether differences in these maps reflect incomplete analysis of retrieved clones or whether the tissues used for the different strategies and subsequent cDNA screenings were not adequate to retrieve all genes.

ACKNOWLEDGMENTS

The 26F3 YAC clone was provided by F. Collins, and the 279F8 and 368C7 YAC clones were provided by D. Lasko (McGill University) and D. Le Paslier (CEPH). This work was supported by the Breast Cancer Research Initiative of the National Cancer Institute of Canada, Fonds de la Recherche en Santé du Québec, Ludwig Institute for Cancer Research, Canadian Genetic Diseases Network (Federal N.C.E. Program), and Endorecherche. J.M.R. and J.S. are Scholars of the Medical Research Council of Canada (M.R.C.). F.D. has a studentship from M.R.C.

REFERENCES

Albertsen, H., Plaetke, R., Ballard, L., Fujimoto, E., Connolly, J., Lawrence, E., Rodriguez, P., Robertson, M., Bradley, P., Milner, B., Fuhrman, D., Marks, A., Sargent, R., Cartwright, P., Matsunami, N., and White, R. (1994a). Genetic mapping of the BRCA1 region on chromosome 17q21. *Am. J. Hum. Genet.* **54:** 516–525.

Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriquez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Mei Sheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994b). A physical map and candidate genes in the BRCA1 region on chromosome 17q12–q21. *Nature Genet.* **7:** 472–479.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Anderson, L. A., Friedman, L., Osborne-Lawrence, L., Lynch, E., Weissenbach, J., Bowcock, A., and King, M.-C. (1993). High density genetic map of the BRCA1 region of chromosome 17q12–q21. *Genomics* **17:** 618–623.

Aoyama, Y., Chan, Y.-L., and Wool, I. G. (1989). The primary structure of rat ribosomal protein L34. *FEBS Lett.* **249:** 119–122.

Bowcock, A., Anderson, L. M., Fridedman, L. S., Black, D. M., Osborne-Lawrence, S., Rowell, S. E., Hall, J. M., Solomon, E., and King, M.-C. (1993). THRA1 and D17S183 flank an interval of <4cM for the breast–ovarian cancer gene (*BRCA1*) on chromosome 17q12–q21. *Am. J. Hum. Genet.* **52:** 718–722.

Brody, L. C., Abel, K. J., Castilla, L. H., Couch, F. J., McKinley, D. R., Yin, G.-Y., Ho, P. P., Merajver, S., Chandrasekharappa, S. C., Xu, J., Cole, J. L., Struewing, J. P., Valdes, J. M., Collins, F. S., and Weber, B. L. (1995). Construction of a transcription map surrounding the BRCA1 locus of human chromosome 17. *Genomics* **25:** 264–273.

Campbell, I. G., Nicolai, H. M., Foulkes, W. D., Senger, G., Stamp, G. W., Allan, G., Boyer, C., Jones, K., Bast, R. C., Solomon, E., Trowsdale, J., and Black, D. M. (1994). A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. *Hum. Mol. Genet.* **3:** 589–594.

Castilla, L. H., Couch, F. J., Erdos, M. R., Hoskins, K. F., Calzone, K., Garber, J. E., Boyd, J., Lubin, M. B., Deshano, M. L., Brody, L. C., Collins, F. S., and Weber, B. L. (1994). Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nature Genet.* **8:** 387–391.

Chamberlain, J. S., Boehnke, M., Frank, T. S., Kiousis, S., Xu, J., Guo, S.-W., Hauser, E. R., Norum, R. A., Helmbold, E. A., Markel, D. S., Keshavarzi, S. M., Jackson, C. E., Calzone, K., Garber, J., Collins, F. S., and Weber, B. L. (1993). BRCA1 maps proximal to D17S579 on chromosome 17q21 by genetic analysis. *Am. J. Hum. Genet.* **52:** 792–798.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Couch, F. J., Castilla, L. H., Xu, J., Abel, K. J., Welcsh, P., King, S. E., Wong, L., Ho, P. P., Merajver, S., Brody, L. C., Yin, G.-Y., Hayes, S. T., Gieser, L. M., Flejter, W. L., Glover, T. W., Friedman, L. S., Lynch, E. D., Meza, J. E., King, M.-C., Law, D. J., Deaven, L., Bowcock, A. M., Collins, F. S., Weber, B. L., and Chandrasekharappa, S. C. (1995). A YAC-, P1-, and cosmid-based physical map of the BRCA1 region on chromosome 17q21. *Genomics* **25:** 264–273.

Cropp, C. S., Nevanlinna, H. A., Pyrhonen, S., Stenman, U.-H., Salmikangas, P., Albertsen, H., White, R., and Callahan, R. (1994). Evidence for involvement of *BRCA1* in sporadic breast carcinomas. *Cancer Res.* **54:** 2548–2551.

Diamond, R. H., Cressman, D. E., Laz, T. M., Abrams, C. S., and Taub, R. (1994). PRL-1, a unique nuclear protein tyrosine phosphatase, affects cell growth. *Mol. Cell. Biol.* **14:** 3752–3762.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D., and Evans, G. A. (1992). A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukemias. *Nature Genet.* **2:** 112–118.

Easton, D. F., Bishop, D. T., Ford, D., Crockford, G. P., and the Breast Cancer Linkage Consortium (1993). Genetic linkage in familial breast and ovarian cancer. *Am. J. Hum. Genet.* **52:** 678–701.

Eccles, D. M., Cranston, G., Steel, C. M., Nakamura, Y., and Leonard, R. C. F. (1990). Allele losses on chromosome 17 in human epithelial ovarian carcinoma. *Oncogene* **5:** 1599–1601.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Friedman, L. S., Ostermeyer, E. A., Szabo, C. I., Dowd, P., Lynch, E. D., Rowell, S. E., and King, M.-C. (1994a). Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nature Genet.* **8:** 399–404.

Friedman, L. S., Ostermeyer, E. A., Lynch, E. D., Szabo, C. I., Anderson, L. A., Dowd, P., Lee, M. K., Rowell, S. E., Boyd, J., and King, M.-C. (1994b). The search for *BRCA1*. *Cancer Res.* **54:** 6374–6382.

Friedman, L. S., Ostermeyer, E. A., Lynch, E. D., Welcsh, P., Szabo, C. I., Meza, J. E., Anderson, L. A., Dowd, P., Lee, M. K., Rowell, S. E., Ellison, E., Boyd, J., and King, M.-C. (1995). 22 genes from chromosome 17q21: Cloning, sequencing, and characterization of mutations in breast cancer families and tumors. *Genomics* **25:** 256–263.

Futreal, P. A., Söderkvist, P., Marks, J. R., Iglehard, J. D., Cochran, C., Barrett, J. C., and Wiseman, R. (1992). Detection of frequent allelic loss on proximal chromosome 17q in sporadic breast carcinoma using microsatellite length polymorphisms. *Cancer Res.* **52:** 2624–2627.

Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J. D., Marks, J., Ballinger, D. G., Barrett, J. C., Skolnick, M. H., Kamb, A., and Wiseman, R. (1994a). *BRCA1* mutations in primary breast and ovarian carcinomas. *Science* **266:** 120–122.

Futreal, P. A., Cochran, C., Rosenthal, J., Miki, Y., Swenson, J., Hobbs, M., Bennett, L. M., Haugen-Strano, A., Marks, J., Barrett, J. C., Tavtigian, S. V., Shattuck-Eidens, D., Kamb, A., Skolnick, M., and Wiseman, R. W. (1994b). Isolation of a diverged homeobox gene, *MOX1*, from the BRCA1 region on 17q21 by solution hybrid capture. *Hum. Mol. Genet.* **3:** 1359–1364.

Goldgar, D. E., Cannon-Albright, L. A., Oliphant, A., Ward, J. H., Linker, G., Swensen, J., Tran, T. D., Fields, P., Uharriet, P., and Skolnick, M. H. (1993). Chromosome 17q linkage studies of 18 Utah breast cancer kindreds. *Am. J. Hum. Genet.* **52:** 743–748.

Goldgar, D. E., Fields, P., Lewis, C. M., Tran, T. D., Cannon-Albright, L. A., Ward, J. H., Swenson, J., and Skolnick, M. H. (1994). A large kindred with 17q-linked breast and ovarian: Genetic, phenotypic and genealogical analysis. *J. Natl. Cancer Inst.* **86:** 200–209.

Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M., and Canaani, E. (1992). The t(4;11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene, related to Drosophila *trithorax*, to the *AF-4* gene. *Cell* **71:** 701–708.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J., Anderson, L., Huey, B., and King, M.-C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q12–q21. *Science* **250:** 1684–1689.

Henikoff, S., and Henikoff, J. G. (1994). Protein family classification based on searching a database of blocks. *Genomics* **19:** 97–107.

Jones, R. S., and Gelbart, W. M. (1993). The *Drosophila* polycomb-group gene *enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol. Cell. Biol.* **13:** 6357–6366.

Jones, K. A., Black, D. M., Brown, M. A., Griffiths, B. L., Nicolai, H. M., Chambers, J. A., Bonjardim, M., Xu, C.-F., Boyd, M., McFarlane, R., Korn, B., Poustka, A., North, M. A., Schalkwyk, L., Lehrach, H., and Solomon, E. (1994). The detailed characterization of a 400 kb cosmid walk in the BRCA1 region: Identification and localization of 10 genes including a dual-specificity phosphatase. *Hum. Mol. Genet.* **3:** 1927–1934.

Kallioniemi, A., Kallioniemi, O.-P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W., and Waldman, F. M. (1994). Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. USA* **91:** 2156–2160.

Kamb, A., Futreal, P. A., Rosenthal, J., Cochran, C., Harshman, K. D., Liu, Q., Phelps, R. S., Tavtigian, S. V., Tran, T., Hussey, C., Bell, R., Miki, Y., Swenson, J., Hobbs, M. R., Marks, J., Bennett, L. M., Barrett, J. C., Wiseman, R. W., and Shattuck-Eidens, D. (1994). Localization of the *VHR* phosphatase gene and its analysis as a candidate for *BRCA1*. *Genomics* **23:** 163–167.

Kelsell, D. P., Black, D. M., Bishop, D. T., and Spurr, N. K. (1993). Genetic analysis of the *BRCA1* region in a large breast/ovarian family: Refinement of the minimal region containing *BRCA1*. *Hum. Mol. Genet.* **2:** 1823–1828.

Lovett, M., Kere, J., and Hinton, L. M. (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88:** 9628–9632.

Luu-The, V., Labrie, C., Simard, J., Lachance, Y., Zhao, H.-F., Couet, J., LeBlanc, G., and Labrie, F. (1990). Structure of two in tandem human 17β-hydroxysteroid dehydrogenase genes. *Mol. Endocrinol.* **4:** 268–275.

Luu-The, V., Labrie, C., Zhao, H.-F., Couet, J., Lachance, Y., Simard, J., Leblanc, G., Cote, J., Berube, D., Gagne, R., and Labrie, F. (1989). Characterization of cDNAs for human estradiol 17β-dehydrogenase and assignment of the gene to chromosome 17: Evidence for two mRNA species with distinct 5′ termini in human placenta. *Mol. Endocrinol.* **3:** 1301–1309.

MacDonald, R. J., Smith, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987). Isolation of RNA using guanidinium salts. *Methods Enzymol.* **152:** 219–234.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, C. J., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266:** 66–71.

Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16:** 1215.

Narod, S. A., Feunteun, J., Lynch, H., Watson, P., Conway, T., Lynch, J., and Lenoir, G. M. (1991). Familial breast–ovarian cancer locus on chromosome 17q12–q23. *Lancet* **338:** 82–83.

Neuhausen, S. L., and Marshall, C. J. (1994). Loss of heterozygosity in familial tumours from three *BRCA1*-linked kindreds. *Cancer Res.* **54:** 6069–6072.

Neuhausen, S. L., Swensen, J., Miki, Y., Liu, Q., Tavtigian, S., Shattuck-Eidens, D., Kamb, A., Hobbs, M. R., Gingrich, J., Shizuya, H., Kim, U.-J., Cochran, C., Futreal, P. A., Wiseman, R. W., Lynch, H. T., Tonin, P., Narod, S., Cannon-Albright, L., Scolnick, M. H., and Goldgar, D. E. (1994). A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene *BRCA1*. *Hum. Mol. Genet.* **3:** 1919–1926.

Nikaido, T., Shimada, K., Shibata, M., Hata, M., Sakamoto, M., Takasaki, Y., Sato, C., Takahashi, T., and Nishida, Y. (1990). Cloning

and nuclectide sequence of cDNA for Ki antigen, a highly conserved nuclear protein detected with sera from patients with systemic lupus erythematosus. *Clin. Exp. Immunol.* **79:** 209–214.

Osborne-Lawrence, S., Welcsh, P. L., Spillman, M., Chandrasekharappa, S. C., Gallardo, T. D., Lovett, M., and Bowcock, A. M. (1995). Direct selection of expressed sequences within a 1-Mb region flanking BRCA1 on human chromosome 17q21. *Genomics* **25:** 248–255.

Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D., and Weissman, S. M. (1991). cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA* **88:** 9623–9627.

Rommens, J. M., Zengerling-Lentes, S., Kerem, B.-S., Melmer, G., Buchwald, M., and Tsui, L.-C. (1989). Physical localization of two DNA markers closely linked to the cystic fibrosis locus by pulsed-field gel electrophoresis. *Am. J. Hum. Genet.* **45:** 932–941.

Rommens, J. M., Lin, B., Hutchinson, G. B., Andrew, S. E., Goldberg, Y. P., Glaves, M. L., Graham, R., Lai, V., McArthur, J., Nasir, J., Theilmann, J., McDonald, H., Kalchman, M., Clarke, L. A., Schappert, K., and Hayden, M. R. (1993). A transcription map of the region containing the Huntington disease gene. *Hum. Mol. Genet.* **2:** 901–907.

Rommens, J. M., Mar, L., McArthur, J., Tsui, L.-C., and Scherer, S. (1994). Towards a transcriptional map of the q21–q22 region of chromosome 7. *In* "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, Eds.), pp. 65–79, Plenum, New York.

Saito, H., Inazawa, J., Saito, S., Kasumi, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K., and Nakamura, Y. (1993). Detailed deletion mapping of chromosome 17q in ovarian and breast cancers: 2-cM region on 17q21.3 often and commonly deleted in tumors. *Cancer Res.* **53:** 3382–3385.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Sato, T., Akiyama, F., Sakamoto, G., Kasumi, F., and Nakamura, Y. (1991). Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* **51:** 5794–5799.

Scherer, S., and Tsui, L.-C. (1991). Cloning and analysis of large DNA molecules. *In* "Advanced Techniques in Chromosome Research" (K. W. Adolph, Ed.), pp. 33–72, Dekker, New York.

Shah, M. B., Guan, X., Einstein, J. R., Matis, S., Xu, Y., Mural, R. J., and Uberbacher, E. C. (1994). "User's Guide to GRAIL and GENQUEST (Sequence Analysis, Gene Assembly and Sequence Comparison Systems) E-mail Servers and XGRAIL (Version 1.2) and XGENQUEST (Version 1.1) Client-Server Systems," available by anonymous ftp to arthur.epm.ornl.gov (128.219.9.76) from directory pub/xgrail or pub/xgenQuest as file Manual.grail-genquest.July 94.

Simard, J., Fuenteun, J., Lenoir, G. M., Tonin, P., Normand, T., Luu-The, V., Vivier, A., Lasko, D., Morgan, K., Rouleau, G., Lynch, H., Labrie, F., and Narod, S. A. (1993). Genetic mapping of the breast–ovarian cancer syndrome to a small interval on chromosome 17q12–21: Exclusion of candidate genes *EDH17B2* and *RARA*. *Hum. Mol. Genet.* **2:** 1193–1199.

Simard, J., Tonin, P., Durocher, F., Morgan, K., Rommens, J., Gingras, S., Samson, C., Leblanc, J.-F., Belanger, C., Dion, F., Liu, Q., Skolnick, M., Goldgar, D., Shattuck-Eidens, D., Labrie, F., and Narod, S. (1994). Common origins of *BRCA1* mutations in Canadian breast and ovarian cancer families. *Nature Genet.* **8:** 392–398.

Smith, S. A., Easton, D. F., Evans, D. G. R., and Ponder, B. A. J. (1992). Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genet.* **2:** 128–131.

Smith, S. A., Dicioccio, R. A., Streuwing, J. P., Easton, D. F., Gallion, H. H., Albertson, H., Mazoyer, S., Johansson, B., Steichen-Gersdorf, E., Stratton, M., Ford, D., Marshall, G., White, R. L., Piver, M. S., and Ponder, B. A. J. (1994). Localization of the breast–ovarian cancer susceptibility gene (*BRCA1*) on 17q12–21 to an interval of <1 cM. *Genes Chrom. Cancer* **10:** 71–74.

Tkachuk, D. C., Kohler, S., and Cleary, M. L. (1992). Involvement of a homologue of *Drosophilia trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71:** 691–700.

Tonin, P., Serova, O., Simard, J., Lenoir, G., Feunteun, J., Morgan, K., Lynch, H., and Narod, S. A. (1994). The gene for hereditary breast–ovarian cancer, *BRCA1*, maps distal to EDHB2 in chromosome region 17q12–q21. *Hum. Mol. Genet.* **3:** 1679–1682.

UnerBacher, E. C., and Mural, R. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88:** 11,261–11,265.

Vogelstein, B., and Kinzler, K. W. (1994). Has the breast cancer gene been found? *Cell* **79:** 1–3.

Zheng, Y., Jung, M. K., and Oakley, B. R. (1991). Gamma-tubulin is present in *Drosophila melanogaster* and *Homo sapiens* and is associated with the centromere. *Cell* **65:** 817–823.

SUBMITTED TO THE AMERICAN JOURNAL OF HUMAN GENETICS

A HIGH INCIDENCE OF BRCA1 MUTATIONS IN 20 BREAST-OVARIAN CANCER FAMILIES

Olga Serova[1], Marco Montagna[1], Delphine Torchard[2], Steven A Narod[3], Patricia Tonin[3], Bakary Sylla[1], Henry T Lynch[4], Jean Feunteun[2], Gilbert M Lenoir[1]

[1] International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon, France; [2] Institute Gustave Roussy, Villejuif, France; [3] Department of Medicine, Division of Medical Genetics, and Department of Human Genetics, McGill University, Montreal, Canada; [4] Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, NE

Correspondence to: Gilbert M Lenoir, International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon, France

RUNNING HEAD: BRCA1 Mutations in Breast-Ovarian Cancer Families

SUMMARY

We have analysed 20 breast/ovarian cancer families, the majority of which show positive evidence of linkage to chromosome 17q12, for germline mutations in the BRCA1 gene. BRCA1 mutations cosegregating with breast and ovarian cancer susceptibility were identified in 16 families, including one family with a case of male breast cancer. Nine of these mutations have not been reported previously. The majority of mutations were found to generate a premature stop codon leading to the formation of a truncated BRCA1 protein of 2% to 88% of the expected normal length. Two mutations altered the RING finger domain. Sequencing of genomic DNA led to the identification of a mutation in the coding region of BRCA1 in twelve families, and cDNA analysis revealed an abnormal or missing BRCA1 transcript in four of the eight remaining families. A total of eight mutations were associated with a reduced quantity of BRCA1 transcript. We were unable to detect BRCA1 mutations in four of the 20 families, but only one of these was clearly linked to BRCA1. It is expected that the majority of clear examples of the breast-ovarian cancer syndrome will be associated with germline mutations in the coding region of BRCA1.

INTRODUCTION

A family history of breast or ovarian cancer is a risk factor for the development of cancer of either type (Lynch et al. 1981). Segregation analysis suggested the existence of an autosomal dominant mutant allele (or alleles), accounting for from 5-10% of breast cancers (Newman et al. 1988; Claus et al. 1991). Breast cancer also appears in families with ovarian cancer (Schildkraut et al. 1989; Go et al. 1983). A gene responsible for hereditary breast and ovarian cancer, BRCA1, was mapped by linkage to chromosome 17 in 1990 (Hall et al. 1990; Narod et al. 1991) and was isolated in 1994 (Miki et al. 1994). This large gene contains 22 coding exons distributed over approximately 100kb of genomic DNA, and produces a protein of 1863 amino acids. A RING finger motif in the amino-terminal portion of the gene suggests that BRCA1 acts as a transcription factor (Miki et al, 1994).

In a recent survey of 63 cancer families, (Shattuck-Eidens et al. 1995) it was found that the majority (86%) of the 37 different BRCA1 mutations were either frameshifts, nonsense mutations or splice site mutations, which presumably lead to premature truncation of the BRCA1 protein. Infrequent regulatory mutations associated with a loss of transcript corresponding to the mutant BRCA1 allele were also seen. Missense mutations were reported in the RING finger domain and elsewhere, but the functional significance of most missense mutations is not yet known.

Although many breast-ovarian cancer families have been found to carry mutations in the coding region of BRCA1, mutations have not been found for the majority of cancer families studied to

date (Castilla et al. 1994; Friedman et al. 1994; Simard et al. 1994; Shattuck-Eidens et al. 1995; Streuwing et al. 1995). It is not yet clear if the absence of mutations was due to the imperfect sensitivity of the screening test used, of if the mutations were outside of the coding region of the gene, or if families included in these studies were due to susceptibility genes other than BRCA1.

Using statistical models, Easton et al. (1995) estimated that there are likely to be two types of BRCA1 families - those with a high penetrance of ovarian cancer (84% by age 70 years) and those with a low penetrance of ovarian cancer (32% by age 70 years). No particular genotype-phenotype correlations have yet been revealed which support this theoretical division. We have studied a panel of 20 breast-ovarian cancer families, each with a high probability of linkage to BRCA1, in order to estimate the sensitivity of direct DNA sequencing and SSCP analysis of cDNA as a two-step method of mutation detection and to identify possible genotype/phenotype correlations.

METHODS

Breast-ovarian cancer families Each of the 20 families contained a minimum of four cases of breast or ovarian cancer, and at least one case of cancer of each type. These families were originally chosen for linkage studies, and therefore blood samples for DNA analysis were available on multiple affected members in each family. Linkage results on several of these families have been reported (Narod et al. 1991; Feunteun et al. 1993; Simard et al. 1994; Torchart et al. 1994; Tonin et al. 1994). All families but one are followed at the Department of Preventive Medicine in Creighton University School of Medicine in Omaha; family 4001 is followed at the Hopital St. Louis, Paris.

The probability of linkage to BRCA1 for each of these families was estimated to be 80% or greater. These probabilities are based on the prior probabilities of linkage, estimated by Narod et al (1995) and the lod scores generated by BRCA1 markers. For example, it is estimated that 81% of families with multiple cases of early-onset breast cancer and a single case of ovarian cancer are linked to BRCA1. A positive lod score will increase the probability of linkage, usually to 90% or higher. The characteristics of the 20 families and the derived probabilities of linkage are presented in table 1. The average age of diagnosis of the 143 breast cancers in these families was 41.8 years, and of the 54 ovarian cancers was 52.6 years. One family contained a case of male breast cancer.

DNA/RNA extraction DNA was extracted from Epstein-Barr virus immortalized lymphoblastoid cell lines using standard methods. RNA isolation was performed with Trizol reagent (Gibco BRL) or with the QuickPrep Micro kit (Pharmacia) according to the instructions provided by the manufacturer.

Linkage analysis To establish the probability that a family was linked to BRCA1, linkage analysis was performed using five polymorphic CA-repeat markers flanking the BRCA1 gene, including the intragenic marker D17S855. The other markers were D17S250, D17S588, D17S579 and THRA1 (Easton et al. 1993). Linkage to BRCA2 was assessed using the chromosome 13 marker AFM238zd9. Multipoint linkage was performed using the LINKAGE program (Lathrop and Lalouel, 1984) under a previously described model (Feunteun et al. 1993). Breast and ovarian cancer susceptibility was modelled as a dominant trait with incomplete penetrance. Affected cases were assigned to liability classes which reflect the age-specific penetrance of breast and ovarian cancer. The risks of sporadic breast cancer (cancer among non-carriers) were derived from the cancer rates in the general population.

Detection of BRCA1 mutations A comprehensive multistep strategy for BRCA1 mutation identification was applied to the 20 families. The first step was the direct sequencing of the entire coding sequence of BRCA1 (with the exception of exon 4 which was analysed by SSCP). Because of the possibility that a BRCA1 mutation may exist outside of the coding region and result in the production of a mutant RNA, if a BRCA1 mutation was not found in the genomic DNA, the cDNA was analysed for mutations by SSCP of RT-PCR products. When abnormal SSCP bands were identified, the corresponding cDNA fragment was analysed by direct sequencing.

In addition, it is also possible that a BRCA1 mutation outside of the coding region may result in the loss of the associated mRNA. An affected individual from each family was therefore studied for the possibility of decreased BRCA1 RNA levels in lymphocytes. This technique is based on the identification of constitutional heterozygosity at one of a number of intragenic BRCA1 polymorphisms. RNA loss is inferred from the preferential presence of sequence corresponding to the normal allele observed in the cDNA sequence ladder (figure 1). This technique is used to identify the class of inferred regulatory mutations, which are characterized by the dominant expression of the normal allele in the lymphocyte DNA.

cDNA synthesis was carried out at 37°C for one hour in a 20 ul reaction volume containing 1 ug of total cellular RNA or 30-50 ng of polyA RNA, 1x buffer (50 mM tris HCl pH 8.3, 75 mM KCl, 3 mM MgCl2, 10 mM DTT), 1 ug random primers (Gibco BRL) 500uM dNTPs (Pharmacia) 10u RNAs (Promega) and 3u MMLV reverse transcriptase (Pharmacia).

47 primer pairs were used to amplify all BRCA1 exons from genomic DNA, and 33 primer pairs were used to amplify BRCA1 cDNA. PCR was performed in 25 ul solutions containing 20-50 ng of genomic DNA or cDNA, 1x PCR buffer, 200 um dNTPs (Pharmacia), 20 pmoles of each primer (Genset), 1 uCi 33P dCTP (Amersham), 1u Taq DNA polymerase (Eurobio). Reactions were performed in a thermal cycler Perkin-Elmer Cetus model 9600 as follows: 94oC for 3 minutes, 35 cycles (94o for 30s, annealing temperature for 30s, 72o for 30s) 72o for 10 minutes. PCR products were denatured at 90oC for 4 minutes and run on Hydrolink MDE gel (AT Biochem) at constant power of 6W for 16h in 0.6xTBE. The gels were dried and autoradiographed for 12-24 hours. The primer sequences used to amplify the exons were from the FTP file at Morgan.Med.UTAH.ed, except for exons 1, 4, 6, 7 and 11, for which those primers published by Friedman et al (1994) were used. The non-coding exon 1 and exon 4 were tested by SSCP analysis.

Direct sequencing of PCR fragments was performed with the sequenase PCR product sequencing kit, using the same primers as for PCR amplification. The samples were migrated on 6% acrylamide gels at 70W. Gels were dried and exposed to X-ray film for 12 hours.

DNA samples from one individual from each family was chosen for sequencing. The chosen individual in each family was observed to carry the haplotype associated with cancer susceptibility in the family.

When a mutation leading to a truncated protein was identified, the remaining exons were not sequenced. When a missense mutation was identified, the entire coding sequence was analysed to rule out the presence of an additional variant. When mutations were found at splice sites, sequencing of RT-PCR products was performed in order to identify the consequence of the BRCA1 mutation at the RNA level. The splice site sequences were screened for mutations by SSCP and confirmed by direct sequencing of PCR fragments amplified from cDNA. Mutations were confirmed by sequencing DNA from two other individuals in the family who were presumed to carry the mutation as well.

Loss of transcript corresponding to the BRCA1 susceptibility allele was performed by comparing the genomic DNA sequence and the cDNA sequence for loss of heterozygosity at polymorphic sites (identified while sequencing the coding region). It was not possible to assay for loss of transcript in three families (family 32, 2619 and 1813) because no heterozygous individuals who carried BRCA1 mutations were identified in these families.

RESULTS

A total of 16 out of the 20 families studied were found to have BRCA1 mutations cosegregating with breast and ovarian cancer susceptibility (table 2). Twelve of these mutations occurred in the coding region, and were identified through the initial sequencing of genomic DNA. Eleven of these were either nonsense or frameshift mutations, and one was a missense mutation. Only one mutation was seen more than once; the deletion of AG dinucleotides at codon 22 was seen in families 2979 and 3079.

The genomic screen also revealed the presence of a splice-site mutation in family 1234. This mutation was detected in the acceptor site of intron 18, and leads to the deletion of a single nucleotide. RT-PCR analysis performed on a carrier from this family revealed that this mutation causes the elimination of exon 19 in the mutant transcript, and premature termination at codon 1732. The levels of normal and mutant BRCA1 transcripts in the lymphocytes of several carriers in family 1234 were studied. The mutant transcript was found to be present at a much lower intensity than the normal allele in each case, and suggests that this mutation is an alteration in RNA metabolism (instability, transcription, transport). This BRCA1 mutation has not been reported previously, but examples of truncating mutations in other genes provoking decreased or absent mRNA levels have been noted (Hamosh 1991; Baserga 1992; Mashima 1992; Stolzenberg et al, 1994).

The RNA sequence was examined in carrier individuals from the seven families for which the genomic sequence was normal. This was done by performing SSCP analysis on the RT-PCR generated cDNA product. One additional mutation was detected by this method. In family 2850 cDNA analysis of lymphocytes revealed that transcripts were missing exon 5 in three carriers. However, no alteration in the splice sites of exon 5, or elsewhere in the genomic sequence was identified. This is similar to the variant lacking exon 3 in the cDNA of carriers reported in one of the Berkeley families, where no genomic change was identified (Friedman et al. 1994). Exon 5 contains three cysteine residues which participate in the formation of the second zinc finger of the RING domain, and the absence of this exon would be expected to have a significant effect on the DNA binding properties of the BRCA1 protein.

Four of the mutations have been seen previously (Shattuck-Eidens et al, 1995) including the mutation in family 1252, in which the RING finger cysteine at position 61 is replaced with glycine (Friedman et al, 1994).

Of the six remaining families for which the genomic and cDNA analyses were normal, two had regulatory mutations inferred from the observation of decreased levels of mRNA. cDNAs of presumed mutation carriers from families 2090 and 3173 showed strong reduction in heterozygosity at the intragenic polymorphic sites,

implying a greatly reduced BRCA1 transcript level (figure 1). No nucleotide changes in genomic DNA were revealed in these two families. In the cDNA of non-carriers, both alleles were found to be of equal intensity. It was not possible to perform this analysis on families 32 and 2619 because no BRCA1 heterozygous individuals were identified.

The relative expression level of the mutant BRCA1 allele was also examined for the other identified mutations, where possible. Four mutations were associated with a reduced relative amount of transcripts, as compared to their normal counterparts (table 2).

It has been suggested that mutations in the 5′ end of the BRCA1 gene may be associated with an increased risk of ovarian cancer (Friedman et al. 1994; Shattuck-Eidens et al. 1995). We saw little evidence to support this hypothesis in our families. In fact, family 1234, which contains nine cases of breast cancer and ten cases of ovarian cancer carries a mutation leading to a truncated protein of nearly full length. There were no significant differences in the age of onset, or of the ratio of ovarian to breast cancers for mutations in the 3′ versus 5′ half of the gene. However, the ratio of cases of ovarian to breast cancer was higher in families with missing or reduced BRCA1 transcript (0.54) compared to families with mutations associated with equal expression of normal and mutant BRCA1 alleles (0.30; $p = 0.14$). Among the eight families with normal RNA levels, there were three families that had mutations leading to termination at codon 39; the ratio of ovarian to breast cancers was higher in these three families (0.53) than in the other five, which carried more distal mutations (0.19; $p = 0.16$).

It was originally reported that families with cases of male breast cancer were unlinked to BRCA1 (Stratton et al. 1994; Narod et al. 1995). Family 2651 (figure 2) is the third example of a BRCA1-linked family which contains a case of male breast cancer (Streuwing et al. 1995, Hogervorst et al. 1995). We were able to demonstrate the presence of the BRCA1 mutation in the male with breast cancer(individual 452). It now appears that the risk of breast cancer is in fact, increased in males carrying BRCA1 mutations.

We failed to find BRCA1 mutations in families 1816, 32, 2619, and 1882. Only family 1816 is clearly linked to BRCA1. The multipoint lod score with D17S579 and THRA1 is 3.62. RNA levels associated with both BRCA1 alleles were normal. Family 32 contains two cases of ovarian cancer and 10 cases of breast cancer (figure 3). It originally appeared that this family was linked to BRCA1. However, in 1994 individual 141 was diagnosed with breast cancer at age 50. She does not share a BRCA1 haplotype of risk with the other affected women, and the lod score is therefore slightly negative. It is possible that she is a sporadic case. Family 2619 (figure 4) contains two cases of breast cancer of relatively late onset (mean age 53.5 years). The relative frequency of ovarian cancer and the presence of

early-onset colon cancer suggests hereditary non-polyposis colon cancer as an alternate diagnosis. Family 1882 had the highest average age of onset of breast cancer in the study (55.8 years), and only a single case of ovarian cancer. There was no evidence from linkage supporting a BRCA1 origin for this family.

All, or nearly all, breast-ovarian cancer families are believed to be linked either to BRCA1 or to BRCA2. Among 145 breast ovarian cancer families studied by the Breast Cancer Linkage Consortium, there was no clear example of a family that was unlinked to both loci (Narod et al, in press). It is expected, therefore, that breast-ovarian cancer families unlinked to BRCA1 should have a high probability of being linked to BRCA2, as was found to be the case in the Consortium data set. However, the lod scores for linkage to BRCA2 for families 32, 2619 and 1882 were -2.50, -1.00, and -0.51, respectively, and do not suggest BRCA2 as the cause of cancer susceptibility in these families. The presence of undetected BRCA1 mutations remains the more plausible explanations.

DISCUSSION

Sensitivity of approach to mutation screening We utilized a two-step approach to BRCA1 mutation screening whereby all coding exons of BRCA1 were sequenced in genomic DNA, and RNA was analyzed by SSCP in families where genomic mutations were not seen. With this approach we have identified BRCA1 mutations in 16 of 20 families (80%) with both breast and ovarian cancer. Mutations were found in 10 of 11 (91%) families with lod scores greater than 0.50; and in 15 of 17 families (88%) where the mean age of diagnosis of breast cancer was less than 50 years. Mutations were not found in four families. One of these has compelling evidence of being linked to BRCA1. The other three kindreds with no BRCA1 mutations were not clearly linked, suggesting the possibility of involvement of another predisposing gene. One previous study used genomic sequencing as a screening method. Simard et al. (1994) detected BRCA1 mutations in 13 of 30 Canadian families with three or more cases of breast or ovarian cancer, but mutations were identified in six of eight families (75%) with two cases of ovarian cancer and two cases of early-onset breast cancer.

Three studies used SSCP as the primary method of mutation screening (Castilla et al. 1994; Friedman et al. 1994; Streuwing et al. 1995). Castilla et al. (1994) detected germline BRCA1 mutations in only eight of 50 familial breast cancer cases. However, only 10 of the 50 families had lod scores of 1.0 or greater to BRCA1. Seven of the eight families positive for mutations contained cases of ovarian cancer. Friedman et al. (1994) also used SSCP and identified BRCA1 mutations in 10 of 20 breast cancer and breast-ovary families with "convincing evidence

of linkage" (lod scores were not given). Streuwing et al. (1995) used SSCP to screen for mutations in 24 families with at least three cases of breast or ovarian cancer (all families contained at least one case of ovarian cancer). Mutations were detected in only 10 of the 24 families, but were found in all four families with lod scores > 0.30.

Hogervorst et al. (1995) screened for BRCA1 mutations using the protein truncation test (PTT). Mutations were identified in only six of 35 patients from breast or breast/ovarian cancer families in this study. However, mutations were found in four of the five families with lod scores greater than 1.0. In our study, it could be anticipated that PTT would detect mutations in 13 families. In six of these 13 families (families 4001, 2651, 2770, 1234, 2090 and 3173) the expression of the mutant allele was significantly reduced. It is not yet clear if this test would perform adequately under these circumstances.

In general, the mutation yield in our study and in previous studies is much higher for families with strong evidence of linkage to BRCA1 (e.g. lod scores > 0.5). The yield is also higher in families with clear-cut evidence of the breast-ovarian cancer syndrome (at least two cases of ovarian cancer and a least two cases of breast cancer before the age of 60). It is estimated that 92% of families satisfying these criteria are attributable to BRCA1 mutations (Narod et al. 1995). It is difficult to evaluate the sensitivity of the SSCP and PTT tests against the method of direct sequencing at present, because it is not clear which negative tests are false negatives. It will be necessary to perform the three tests on a common set of families if comparisons of sensitivity are to be meaningful.

Mutation types The majority of mutations in our study (81%) resulted in truncated BRCA1 proteins, as reported by Shattuck-Eidens et al. (1995). However, we found a different distribution of mutation subtypes within this class, notably a relative excess of nonsense and splice-site mutations (table 3). These differences are likely to be the result of the methods used for screening in the different studies. In our study, both genomic DNA and cDNA were analysed. The approach of direct DNA sequencing is expected to lead to a greater rate of detection of single base-pair substitutions (such as nonsense mutations) than is SSCP. Few previous studies have looked for regulatory mutations, or for other non-coding mutations generating an unstable or mutant BRCA1 transcript. In addition, our series included families selected for both breast and ovarian cancer.

Genotype/Phenoytpe Correlations Overall, we did not find a significant influence of the position of the BRCA1 mutation on the appearance of the breast and ovarian cancers in these families, either in terms of the ratio of breast to ovarian cancers, or in terms of the average age of onset of these

cancers. In particular, we were unable to confirm the speculation of Shattuck-Eidens et al (1995) that 5′ mutations confer a greater risk of ovarian cancer. Of the two families with the greatest number of ovarian cancers in this series, one had a mutation in intron 18 (near the 3′ end) and the other did not have any mutation identified in the BRCA1 coding region. Among the 14 informative families with BRCA1 mutations, we found a reduction in RNA levels in six families, and equal expression of the two alleles in eight. There is a suggestion that the risk of ovarian cancer is greater in families with mutations associated with reduced RNA levels. However, it should be noted that within this subgroup was a family with 3 ovarian and 14 breast cancers, and another with 10 ovarian and 9 breast cancers. In the subgroup of eight families with normal BRCA1 RNA levels, the penetrance of ovarian cancer was higher in three families with extreme 5′ mutations than in the other five. These preliminary observations require confirmation. It may be that other factors, genetic and non-genetic, have greater importance in determining cancer penetrance than the allelic variation of the BRCA1 gene itself.

Issues in Genetic Counselling It is possible to offer predictive testing in some families for the breast-ovarian cancer syndrome using linkage analysis alone (Lynch et al, 1994), but the process is greatly facilitated when direct mutation detection is also available. We have taken a comprehensive approach to risk assessment which includes both linkage analysis and direct mutation analysis detection. In several cases, the combined approach has allowed us to provide evaluations when a single approach was insufficient. We have provided risk assessments based on DNA analysis to over 200 individuals from 10 of these families. Haplotype and mutation analyses are done in different laboratories, and the co-segregation of the identified mutation with the associated haplotype is confirmed prior to risk communication. Although we were unable to detect a BRCA1 mutation in family 1816 the multipoint lod score for linkage to chromosome 17q markers in this family is greater than 3.0, and the probability of BRCA1 linkage is estimated to be above 99%. We have provided DNA-based risk estimates to over 100 individuals in this single family based on linkage analysis alone.

The linkage results for family 2651 were difficult at first to interpret. The affected persons in the left hand branch of the pedigree (individuals 67, 60, 107 and 120; figure 2) did not share an affected haplotype with the five affected women in the middle branch (individuals 1, 6, 29, 45, and 39). Although the segregation of the (3,5,3,6,2) haplotype in the latter five women was consistent with a BRCA1 mutation inherited from individual 69, we were unwilling to provide risk assesments based on this assumption. The finding of the 5 base-pair deletion in exon 11 segregating with this haplotype confirmed this pattern of inheritance. The family history of individual 69 was then extended, and additional cases of cancer, including a case of male breast cancer in individual 452, were revealed. Genetic risks have now been given to family members based on the combined

mutation and linkage data.

The principal considerations for offering predictive testing are the specificity and sensitivity of the test employed. In the absence of strong evidence of linkage, a negative mutation test is not useful; i.e., it is not possible to reassure individuals that they are not at elevated risk, because of the possibility that a mutation may have been missed, and because of genetic heterogeneity. However, once a mutation is identified in an affected individual it is possible to offer accurate predictive testing to unaffected women in a family. Currently, we feel that there is sufficient evidence to conclude that a mutation that leads to premature protein truncation is a true positive, and we counsel families accordingly. In contrast, we are reluctant to offer counselling based solely on the presence of a missense or an inferred regulatory mutation, and we require confirmatory linkage evidence. Our data do not support the hypotheses that 5′ mutations carry a greater risk of ovarian cancer than do 3′ mutations. Because we believe there is no convincing evidence yet for genotype/phenotype correlations, we feel that it is premature to modify our risk estimates based on the position of the BRCA1 mutations.

Table 1. Breast-Ovary Cancer Families used in the Analysis

Family Number	Number of Breast Cancers	Average Age of Diagnosis	Number of Ovarian Cancers	Average Age of Diagnosis	LOD Score	Probability of Linkage to BRCA1
32	10	45.2	2	68.0	-0.41	80
1086	11	41.4	1	65	0.38	91
1234	9	43.7	10	54.1	1.00	99
1252	7	37.6	1	84	0.73	96
1813	5	36.2	1	51	0.54	94
1816	14	40.1	11	50.4	3.62	99
1882	6	55.8	1	73	-0.03	80
1973	5	35.6	1	53.0	0.52	93
2090	7	32.3	2	41.5	0.03	92
2619	2	53.5	4	59.5	0.16	94
2651	5	45.0	2	52.0	-0.05	91
2770	14	39.6	3	47.7	1.56	99
2775	10	43.7	1	47	1.46	99
2850	16	41.6	2	46.5	1.26	99
2944	5	35.4	2	47	0.80	99
2979	2	38.4	2	47	1.15	99
3079	3	49.7	5	53.6	0.45	97
3173	1	51	4	54.5	0.28	96
3300	4	40.5	1	47	0.50	93
4001	7	40.7	2	43.0	0.30	96

Footnote to table 1. The two-point lod scores are based on linkage to the intragenic chromosome 17q markers D17S855 under a model previously described (Feunteun et al, 1993), except for family 1816 which is based on the D17S579 and THRA1 markers. The probabilities of linkage are based on the estimates by Narod et al (1995), and incorporate the lod score information. Family 2651 contains a case of male breast cancer.

Table 2. BRCA1 Mutations in 20 Families Studied

Family Number	Type	Exon	Codon	Nucleotide	Mutation	Loss of Transcript
2979	FS	2	22	185	del AG - ter 39	No
3079	FS	2	22	185	del AG - ter 39	No
2775	FS	2	23	188	del 11 - ter39	No
1252	MS	5	61	300	Cys-Gly (T-G)	No
2850	SP	5	-	-	exon 5 missing in transcript	
1086	FS	11	266	916	del TT - ter 285	No
4001	NS	11	526	1695	Gln-Ter (C-T)	Yes
2944		11	908	2841	Glu-Ter (G-T)	No
3300	FS	11	1160	3598	del 11 - ter 1166	No
2651	FS	11	1234	3819	del 5 - ter 1242	Yes
2770	NS	16	1563	4808	Tyr-Ter (C-G)	Yes
1234	SP	Intron 18	-	- -	del A - del exon 19 - ter 1732	Yes
1973	FS	20	1756	5382	ins C - ter 1829	No
1813	NS	24	1835	5622	Arg-Ter (C-T)	Not done
2090	IR	-	-	-	-	Yes
3173	IR	-	-	-	-	Yes

FS Frameshift; MS Missense; NS Nonsense; IR Inferred Regulatory Transcript analysis was not done for family 1813 because RNA was unavailable from an informative (heterozygous) individual carrying the BRCA1 mutation in this family.

Table 3. Frequency of Different Types of BRCA1 Mutations

Type of Mutation	Mutations Found (%)	Mutations Reported (%)
Frameshift	7 (43.8)	45 (71.4)
Nonsense	4 (25.0)	6 (9.5)
Splice site	2 (12.5)	2 (3.2)
Missense	1 (6.3)	9 (14.3)
Inferred Regulatory	2 (12.5)	1 (1.6)
Total	16	63

Footnote: Mutations reported are from Shattuck-Eidens et al (1995).

REFERENCES

Baserga SJ and Benz EJ (1992). Beta-globin nonsense mutation:deficieint accumulation of mRNA occurs despite normal cytoplasmic stability. Proc Natl Acad Sci USA 89, 2935-2939.

Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE et al (1994) Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. Nature Genetics 8, 387 -391.

Claus EB, Risch N, Thompson WD (1991). Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am J Hum Genet. 48:232-242.

Easton DF, Bishop DT, Ford D, Crockford GP and the Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer - results from 214 families. Amer J Hum Genet 52:678-701

Easton DF, Ford C, Bishop TD and the Breast Cancer LInkage Consortium (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 56:265-271.

Feunteun J, Narod SA, Lynch HT, Watson P, Conway P, Lynch J, Parboosingh J et al. (1993) A breast cancer susceptibility gene maps to chromosome 17q21. Am J Hum Genet 52:736-743

Friedman LS, Ostermeyer EA, Szabo CI, Dowd P, Lynch ED, Rowell SE, King M-C (1994) Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nature Genetics 8; 399-404.

Go RCP, King M-C, Bailey-Wilson J, Elston RC, Lynch HT (1983). Genetic epidemiology of bresat cancer and associated cancers in high risk families. I. Segregation analysis. J Natl Cancer Inst 71:455-461.

Hall JM, Lee MK, Newman B, Morrow J, Anderson A, Huey B, King M-C (1990) Linkage analysis of early onset familial breast cancer to chromosome 17q21. Science 250:1684-1689

Hamosh A, Trapnell BC, Zeitlin PL, Montrose-Rafizadeh C, Rosenstein BJ, Crystal RG, Cutting GR (1991). Severe deficiency of cystic fibrosis transmembrane conductance regulator messenger RNA carrying nonsense R553X and W1316X in respiratory epithelial

cells of patients with cystic fibrosis. J Clin Invest 88:1880-1885.

Hogervorst FBL, Cornelis RS, Bout M, van Vliet MV, Oosterwijk JC< Olmer R et al. (1995) Rapid detection of BRCA1 mutations by the protein truncation test. Nature Genetics 10:208-212.

Lathrop GM, Lalouel JM (1984) Easy calculation of lod scores and genetic risks on small computers. Am J Hum Genet 36:460-465

Lynch HT (1981) Genetics and breast cancer. Van Nostrand Rheinhold, NY.

Lynch HT, Watson P, Conway TA, Lynch JF, Slominski-Carter SM, Narod SA et al. (1993) Pilot study of DNA screening for breast/ovarian cancer susceptibility based on linked markers. Arch Int Med 153:1979-1987.

Mashima Y, Murakami A, Weleber RG, Kennaway NG, Clarke L, Shiono T, Inana G (1992) Nonsense codon mutations of the ornithine aminotransferase gene with decreased levels of mutant mRNA in gyrate atrophy. Am J Hum Genet 51:81-91.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal FA, Harshman K, Tavrigian S et a. (1994) A strong candidate gene for the breats and ovarian cancer susceptibility gene BRCA1. Science 266: 66-71.

Narod SA, Feunteun M, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-23. Lancet 338:82-83

Narod SA, Ford D, Devilee P, Barkardottir RB, Lynch HT, Smith SA et al (1995). An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Amer J Hum Genet 56:254-264.

Narod SA, Ford D, Devilee P, Barkardottir R, Eyfjord J, Lenoir G et al (1995) Genetic heterogeneity of breast-ovarian cancer revisited. Am J Hum Genet (in press).

Newman B, Austin MA, Lee M, King MC (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high risk families. Proc. Natl. Acad. Sci. USA 85:3044-3048.

Schildkraut JM, Risch N, Thompson WD (1989). Evaluating genetic associations among ovarian, breast and endometrial cancer: evidence for a breast/ovarian cancer relationship. Amer J Human Genet 45:521-529.

Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod SA, Couch F et al (1995) A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. J Amer Medic Assoc 273:535-541.

Simard J, Feunteun J, Lenoir GM, Tonin P, Normand T, The VL, Vivier A et al (1993) Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12-21: exclusion of candidate genes EDH17B2 and RARA. Hum Molec Genet 2:1193-1199

Simard J, Tonin P, Durocher F, Morgan K, Rommens J, Gingras S et al. (1994) Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. Nature Genetics 8: 392-398.

Stolzenberg MC, Brugieres L, Gardes M, Dessarps-Freichey F, Chompret A, Bressac B, Lenoir G et al. (1994) Germline exclusion of a single p53 allele by premature termination of translation in a Li-Fraumen syndrome family Oncogene 9: 2799-2804.

Stratton MR, Ford D, Neuhasen S, Seal, Wooster R, Friedman LS, King M-C et al.(1994) Familial male breast cancer is not linked to the BRCA1 locus on chromosome 17q. Nature Genetics 7:103-107.

Streuwing JP, Brody LC, Erdos MR, Kase RG, Giambarresi TR, Smith SA et al. (1995) Detection of eight BRCA1 mutations in 10 breast/ovarian cancer families, including 1 family with male breast cancer. Amer J Hum Genet 57: 1-7.

Torchart D, Serova O, Lenoir G, Langbein C, Narod SA, Janin N et al (1994). A missense mutation in the rod domain of keratin 9 segregates in a pedigree with hereditary epidermolytic palmoplantar keratoderma and with breast/ovarian cancers. Nature Genetics 6:106-110.

Tonin P, Serova O, Simard J, Lenoir G, Feunteun J, Morgan K et al. (1994) The gene for hereditary breast-ovarian, BRCA1 maps distal to EDH17B2 in chromosome region 17q12-21. Hum Molec Genet 3:1679-1682.

LEGENDS FOR FIGURES

Figure 1. Analysis of the relative amount of BRCA1 transcripts produced from mutant and normal alleles in lymphocytes of BRCA1 mutation carriers in family 2090. PCR fragment N of exon 11 (Friedman et al, 1994) was amplified from lymphocyte cDNA and genomic DNA was sequenced and the intensities of the two alleles A and G of the single base polymorphisms at position 3667 was compared. The A allele was found to segregate with cancer susceptibility in this family. 1 and 2: sequences from cDNA in two different carrier individuals; 3 and 4: corresponding genomic sequences for these individuals; 5: cDNA sequence of noncarrier in this family heterozygous for the polymorphism. Note the reduced intensity of the G allele (arrow) in the cDNA from carriers, but not in the genomic, or non-carrier DNA.

Figure 2. Pedigree of family 2651. Black circles indicate affected women. Diagonal slash indicates deceased. Individual identification numbers appear directly below the symbols. Br = breast cancer, Ov = ovarian cancer, co = colon cancer, Gb = cancer of gall bladder, Csu = cancer, site unknown, Sto = stomach cancer, Pro = prostate cancer, En = endometrial cancer, Bt = Brain tumour, Lu = lung cancer. The numbers following these abbreviations indicate ages of diagnosis, where known. The numbers arranged vertically below the individual symbols indicate the marker alleles arranged into haplotypes. Haplotypes in brackets are inferred. Marker alleles separated by a comma cannot be phased. A dash in the place of a marker typing indicates missing information. The haplotype in the vertical rectangle is that observed to segregate with the BRCA1 mutation. The presence of the mutation is indicated by an asterisk, the N refers to the absence of this mutation.

Figure 3. Pedigree of family 32. Symbols as in figure 1. Eso = esophageal cancer; Lip = cancer of the lip.

Figure 4. Pedigree of family 2619. Symbols as in figure 1.

ACKNOWLEDGEMENTS

We thank the family members and the physicians who contributed to this work. We thank MF Lavoue, C Bonnardel, S Delaloge, T Conway, J Lynch, P Watson and S Slominski for their expert assistance. This work was supported by grants from the Ligue Nationale Contre le Cancer, the Canadian Genetics Diseases Network and the Department of the Army (US) Grant number DAMD17-94-J-4299 and the National Institutes of Health grants 1RO1 CA 63678-01. OS is a fellow of the fondation Merieux. MM is a fellow of the Associazione Italiana per la Ricerca sur Cancer (AIRC) and is a recipient of an IARC special training award. SN is a fellow of the Fonds de la Recherche en Sante du Quebec and PT is supported by the Medical Research Council of Canada.

A Collaborative Survey of 80 Mutations in the *BRCA1* Breast and Ovarian Cancer Susceptibility Gene

Implications for Presymptomatic Testing and Screening

Donna Shattuck-Eidens, PhD; Melody McClure; Jacques Simard, PhD; Fernand Labrie, MD, PhD; Steve Narod, MD; Fergus Couch, PhD; Kent Hoskins, MD; Barbara Weber, MD; Lucio Castilla, PhD; Mike Erdos; Lawrence Brody, PhD; Lori Friedman; Elizabeth Ostermeyer; Csilla Szabo, PhD; Mary-Claire King, PhD; Suresh Jhanwar, PhD; Kenneth Offit, MD; Larry Norton, MD; Teresa Gilewski, MD; Mathew Lubin, MD; Michael Osborne, MD; Donald Black, PhD; Marie Boyd; Michael Steel, MD; Sue Ingles; Robert Haile, PhD; Annika Lindblom, MD; Hakan Olsson, MD, PhD; Ake Borg, PhD; D. Timothy Bishop, PhD; Ellen Solomon, PhD; Paolo Radice, MD; Giovanbattista Spatti, MD; Simon Gayther, PhD; Bruce Ponder, MD, PhD; William Warren; Mike Stratton, MD, PhD; Qingyun Liu, PhD; Frank Fujimura, PhD; Cathryn Lewis, PhD; Mark H. Skolnick, PhD; David E. Goldgar, PhD

Objectives.—To report the initial experience of an international group of investigators in identifying mutations in the *BRCA1* breast and ovarian cancer susceptibility gene, to assess the spectrum of such mutations in samples from patients with different family histories of cancer, and to determine the frequency of recurrent mutations.

Design.—Nine laboratories in North America and the United Kingdom tested for *BRCA1* mutations in DNA samples obtained from a total of 372 unrelated patients with breast or ovarian cancer largely chosen from high-risk families. Three of these laboratories also analyzed a total of 714 additional samples from breast or ovarian cancer cases, including 557 unselected for family history, for two specific mutations that had been found to recur in familial samples.

Participants.—A total of 1086 women with either breast or ovarian cancer.

Main Outcome Measure.—The detection of sequence variation in patients' DNA samples that is not found in sets of control samples.

Results.—*BRCA1* mutations have now been identified in a total of 80 patient samples. Thirty-eight distinct mutations were found among 63 mutations identified through a complete screen of the *BRCA1* gene. Three specific mutations appeared relatively common, occurring eight, seven, and five times, respectively. When specific tests for the two most common mutations were performed in larger sets of samples, they were found in 17 additional patients. Mutations predicted to result in a truncated protein accounted for 86% of the mutations detected by complete screening.

Conclusions.—The high frequency of protein-terminating mutations and the observation of many recurrent mutations found in a diverse set of samples could lead to a relatively simple diagnostic test for *BRCA1* mutations. More data must be accumulated to address specifically the sensitivity and specificity of such a diagnostic testing procedure and to better estimate the age-specific risk for breast and ovarian cancer associated with such mutations.

(*JAMA*. 1995;273:535-541)

A PROMISING approach for reducing the high incidence and mortality associated with breast and ovarian cancer lies in the early detection of women at high risk. These women, once identified, can be targeted for more aggressive prevention programs. Because the strongest known epidemiologic risk factor for breast cancer is a positive family history[1] and because studies of breast cancer patients and their relatives consistently find statistical evidence for involvement of an autosomal dominant gene,[2,3] the identification of specific genetic effects have long been the focus of efforts to identify women at high risk. Six months ago, a specific gene, *BRCA1*,[4] which confers greatly increased susceptibility to breast and ovarian cancer, was isolated,[5] creating great interest among oncologists and among women with a personal or family history of breast cancer. The focus of this article is on issues in diagnostic testing specifically related to *BRCA1*, while the general issues of risk assessment and counseling for individuals with a family history of breast cancer are addressed elsewhere in this issue.[6]

BRCA1 is a large gene, containing 5592 nucleotides spread over approximately 100 000 bases of genomic DNA; it is composed of 22 coding exons producing a protein of 1863 amino acids. Much of *BRCA1* shows no homology to other known genes, with the exception of a 126 nucleotide sequence at the amino terminus, which encodes a RING finger motif (a configuration characteristic of the *RING-1* gene[7]), a motif found in other proteins that interact with nucleic acid and/or form protein complexes. This may suggest a role for *BRCA1* in DNA transcription.

Mutations in *BRCA1* result from different types of DNA alteration. Two classes of mutations result from a single nucleotide substitution. If the substitution changes a single amino acid but does

A complete list of author affiliations appears at the end of this article. Drs Skolnick, Lewis, and Goldgar have stock options in Myriad Genetics, which may offer diagnostic testing for *BRCA1* in the future. Dr Skolnick is vice president of Myriad Genetics.

Reprint requests to Department of Medical Informatics, University of Utah School of Medicine, 391 Chipeta Way, Suite D2, Salt Lake City, UT 84108 (Dr Goldgar).

not affect the remainder of the protein translation, it is termed a *missense* mutation. *Nonsense* mutations occur when the nucleotide substitution produces a stop codon (TGA, TAA, or TAG), and translation of the protein is terminated at that point. *Frameshift* mutations occur when one or more nucleotides are either inserted or deleted. If the number of bases inserted or deleted is not divisible by three, a change in the reading frame alters the remainder of the translation of the protein being altered; most often a stop signal is encountered prematurely. Another class of mutation, intron/exon splice-site mutations, can result from either single base changes or the insertion or deletion of one or more nucleotides in the intronic sequence. Splice-site mutations cause abnormal inclusion or exclusion of DNA in the coding sequence, resulting in an aberrant protein. A final class of mutation occurs when a mutation in a gene's regulatory region causes reduction or loss of protein synthesis from the mutant chromosome. Such regulatory mutations usually occur outside of the coding sequence of a gene and can be either nucleotide substitutions or insertion/deletion events.

There are several methods that can be used to detect mutations. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing, can detect sequence variation. For a gene as large as *BRCA1*, manual sequencing is labor intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed using this method. Another approach is the single-strand conformation polymorphism (SSCP) assay.[8] This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 base pairs, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCP make it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments that have shifted mobility on SSCP gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on detection of mismatches between the two complementary DNA (cDNA) strands include clamped denaturing gel electrophoresis[9] and heteroduplex analysis.[10] None of the methods just mentioned will detect large deletions, duplications, or insertions, which might affect one or more complete exons or even the entire gene, nor will they detect a regulatory mutation, which affects transcription or translation of the protein. Other methods that might detect these classes of mutations, such as the protein truncation assay,[11] detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe.[12] Once a mutation is known, an allele-specific detection approach, such as allele-specific oligonucleotide (ASO) hybridization, can be used to screen large numbers of other samples rapidly for that same mutation.

The utility of any screening test, genetic or otherwise, depends on three basic factors: the prior probability of having the condition of interest, the sensitivity of the test (the chance that a person with the condition will have a positive test result), and the specificity of the test (the chance that a person without the condition will have a negative test result). For *BRCA1*, we must consider the probability that a woman diagnosed with breast cancer at a particular age or with a specific family history harbors a mutation in *BRCA1*; that a mutation will be detected using a given screening technique; and that a benign sequence variant will be interpreted as a deleterious *BRCA1* mutation.

Much of what is currently known about the risks of cancer due to *BRCA1* susceptibility comes from analyses of data collected as part of an international effort to study breast cancer families for linkage to the *BRCA1* region of chromosome 17q.[13-15] The probabilities of various constellations of personal and family history being attributable to *BRCA1* can be calculated from statistical models of the penetrance of the *BRCA1* locus. These penetrance estimates have been obtained from analyses of large numbers of *BRCA1*-linked kindreds and on analysis of population-based studies of familial breast and ovarian cancer. The probability of a woman with breast or ovarian cancer bearing a *BRCA1* mutation based on such modeling is given in Table 1. These risks may be overestimates since they depend in part on results obtained from high-risk families suitable for linkage analysis. Now that the *BRCA1* gene has been isolated, more precise estimates of the proportion of women diagnosed with breast and ovarian cancer at a given age and with a given family history that is attributable to *BRCA1* mutations can be determined.

In an analysis of 33 families with at least four cases of breast or ovarian cancer and with sufficient linkage evidence to yield a posterior probability of 17q linkage of greater than 0.9, Easton et al[16] found significant statistical evidence for heterogeneity in the risk of ovarian cancer; it was estimated that 89% of families had a relatively low risk of ovarian cancer (26% by age 70 years), and the remaining 11% of families had a much higher risk (85%). Interestingly, no such heterogeneity was observed in breast cancer risk. The presumed explanation for this finding is allelic heterogeneity; that is, different mutations may confer inherently different risks of ovarian cancer. Now that *BRCA1* has been isolated, this hypothesis can be tested directly. Screening and prevention recommendations would vary according to the age-specific breast cancer risk and the ovarian cancer risk associated with a specific mutation.

Table 1.—Estimated Prior Probabilities of a Particular Constellation of Cases Being Due to the *BRCA1* Gene

Criterion, Cancer Diagnosed, Age (y)	Prior Probability Case Has a *BRCA1* Mutation*
Single affected	
Breast <30	0.12
Breast <40	0.06
Breast 40-49	0.03
Ovarian <50	0.07
Sister pairs	
Breast <40, breast<40	0.37
Breast 40-49, breast 40-49	0.20
Breast <50, ovarian <50	0.46
Ovarian <50, ovarian <50	0.61
Families	
Breast only, ≥3 cases diagnosed <50	0.40
≥2 breast with ≥1 ovarian	0.82
≥2 breast with ≥2 ovarian	0.91

*Estimated using the model of Easton et al[13] for penetrance of *BRCA1* and the estimates of heterogeneity reported in Easton et al.[16]

In the current article, we use data gathered from the preliminary experience of many groups of investigators in screening the *BRCA1* gene to begin answering some of these questions. These investigators screened samples ascertained from a wide variety of sources and used a variety of techniques to detect *BRCA1* mutations. Included in this article are the original *BRCA1* mutations reported by Miki et al[5] and Futreal et al,[17] the mutations found in 30 Canadian families by Simard et al,[18] the analysis of Michigan families by Castilla et al,[19] and the analysis of Friedman et al[20] of the 20 families studied by the Berkeley (Calif) group. In addition, subsequent mutations by these and other groups are reported herein for the first time. The samples analyzed range from those shown with high probability by linkage analysis to segregate *BRCA1* mutations, to small families with two or three affected individuals and no linkage data, to samples from early-onset cases unselected for family history. Our goal in this initial review is to give an early view of what the mutation profile might look like and what it implies for the complexity of an eventual diagnostic test.

Table 2.—Sets of DNA Samples Screened for Mutations in *BRCA1*

Sample Set, Reference or Site*	Description of Samples†	Screening Method‡	No. of Samples Screened	No. of Mutations Found to Date§
Utah-1, Miki et al[5]	Linked families	SEQ	8	5
Utah-2, MGI	Breast/ovarian families	SEQ	10	2
Montreal, Quebec, Simard et al[18]	Breast/ovarian families	SEQ	30	13
Michigan, Castilla et al[19]	Breast and breast/ovarian families	SSCP	39	6 (3)
Berkeley, Calif, Friedman et al[20]	Linked families	SSCP	20	9
MSKCC, MGI	Breast and breast/ovarian families	SEQ	14	2
MSKCC/Utah-1, MGI	Early-onset breast cases	SEQ	24	1
Strang Center, New York, NY (MGI)	Breast and breast/ovarian families	SEQ	12	4
Stockholm, Sweden, MGI	Breast and breast/ovarian families	SEQ	15	3
University of Lund, Sweden (University of California, Berkeley)	Breast and breast/ovarian families	SSCP	22	3
Glasgow-1, Beatson Institute, Glasgow, Scotland	Breast and breast/ovarian families	SEQ/SSCP/ASO	49	2 (1)
Institute of Cancer Research, Sutton, England	Breast and breast/ovarian families	SEQ	30	2 (1)
CRC Group, Cambridge, England	Ovarian families	SSCP/SEQ/HA	30	3
Imperial Cancer Research Fund, London, England	Breast and breast/ovarian families	SEQ	4	2
USC-1, MGI	Bilateral breast cancer proband, high risk	SEQ	7	2
Tumor-1, Futreal et al[17]	Ovarian tumors	SEQ	12	1
Tumor-2, Futreal et al[17]	Early-onset breast tumors	SEQ	32	3
Tumor-3, MGI	Early-onset breast tumors	SEQ	14	1
USC-2, MGI	Bilateral breast cancer diagnosis at <50 y and one additional relative with breast cancer	ASO	59	5
Glasgow-2 Beatson Institute, Glasgow, Scotland	Premenopausal breast cases	ASO	98	3
Glasgow-3 Beatson Institute, Glasgow, Scotland	Unselected ovarian cases	ASO	100	2
MSKCC/Utah-2, MGI	Early-onset breast cases	ASO	109	3
CRC-ovarian, CRC Group, Cambridge, England	Unselected ovarian cases	ASO	250	1
Milan, Italy, MGI	Breast, ovarian, and breast/ovarian families	ASO	50	2

*Reference for published mutations or laboratory in which samples were screened for unpublished data. MGI indicates Myriad Genetics, Salt Lake City, Utah; MSKCC, Memorial Sloan-Kettering Cancer Center, New York, NY; and USC, University of Southern California, Los Angeles.

†Most sample groups contained a heterogenous mixture of samples. The most representative description of each set is given.

‡SEQ indicates direct sequencing of polymerase chain reaction products; SSCP, single-strand conformation polymorphism; ASO, allele-specific oligonucleotide; and HA, heteroduplex analysis.

§Numbers in parentheses indicate mutations identified by laboratory that were independently identified by another laboratory listed in table.

METHODS

Nine laboratories in North America and the United Kingdom screened a total of 372 samples for *BRCA1* mutations, derived from patients with breast or ovarian cancer, drawn primarily from high-risk families. Three of these laboratories tested for two recurrent mutations found in the familial samples, using 714 additional samples derived from breast or ovarian cancer cases, including 557 unselected for family history.

The DNA samples that were screened for *BRCA1* mutations were extracted from blood or tumor samples from patients with breast or ovarian cancer (or known carriers by haplotype analysis) who were participating in research studies on the genetics of breast cancer. All subjects signed appropriate informed consent forms. Table 2 details the number of samples, ascertainment criteria, and screening method for each set of samples screened. Although the original mutations described by Miki et al[5] were detected through screening of cDNA, 25 pairs of intronic polymerase chain reaction primers were used to amplify the complete coding sequence and splice junctions from genomic DNA for the majority of the remaining samples. Updated primer information is published in Castilla et al[19] and Friedman et al.[20] All putative mutations were presumed to be resequenced to verify the original observation. Where possible, DNA sequence variations were tested for cosegregation with breast or ovarian cancer in the family. Further evidence of a causal role of a sequence variant in cancer was provided in some cases by proving the absence of the putative mutation in a set of control individuals. Screening for specific previously identified mutations in large sets of selected samples was performed using ASO hybridization.

In a preliminary attempt to test if alterations in certain domains of *BRCA1* were more likely to be associated with a particularly high proportion of cases of ovarian cancer in these families, we divided the samples in which mutations were found as follows:

1. All proband samples from families with a minimum of three cases of ovarian cancer in which the number of breast cancer cases was no more than twice the number of ovarian cancer cases were classified as having a high proportion of ovarian cancer cases.

2. Those samples from families with at least three cases of breast cancer and in which the number of breast cancer cases was more than twice the number of ovarian cancer cases were denoted as having a low proportion of ovarian cancer cases.

3. All other samples were classified as having an indeterminate or unknown proportion of ovarian cancer cases.

These classifications roughly correspond to a classification based on the posterior probability of being a type I or type II family in the model of Easton et al.[16]

Tests that the observed mutations were distributed uniformly across the coding region of *BRCA1* and comparisons of the locations of mutations according to the classification of high and low proportion of ovarian cancer cases just described were performed using the one-sampleandtwo-sampleKolmogorov–Smirnov tests, respectively.

RESULTS

We distinguish mutations identified by analyzing the entire coding region of *BRCA1* from those in which screening is targeted to specific regions of the gene or specific mutations via ASO analysis. Table 3 describes each of the 63 mutations found by screening the entire *BRCA1* coding sequence plus the 21

Table 3.—Mutations Identified by Complete Screening of the Coding Region in Genomic DNA or Complementary DNA Samples From Material Described in Table 2 (Mutations are Listed in Order of Location From the 5′ End to the 3′ End of *BRCA1*)

Sample Set*	Family	No. of Cases†		Mutation Description			
		Breast	Ovarian	Type‡	Exon	Codon	Mutation§
Utah-1	2035	10	1	IR	...	...	IR
Montreal, Quebec	180	2	2	FS	2	23	*185 del AG→*ter 39
Montreal, Quebec	235	4	2	FS	2	23	*185 del AG→*ter 39
Montreal, Quebec	253	1	3	FS	2	23	*185 del AG→*ter 39
Montreal, Quebec	255	6	7	FS	2	23	*185 del AG→*ter 39
CRC Group, Cambridge, England	OV010	1	3	FS	2	23	*185 del AG→*ter 39
MSKCC	43-04	3	0	FS	2	23	*185 del AG→*ter 39
USC-1	008	2	1	FS	2	23	*185 del AG→*ter 39
ICRF	BOV3**	13	2	FS	2	24	*188 del 11→*ter 39
Utah-1	1901	10	1	FS	5	61	Cys 61 Gly
Berkeley, Calif	4	10	0	MS	5	61	Cys 61 Gly
Michigan	84	5	0	MS	5	64	Cys 64 Gly
Berkeley, Calif	27	8	5	SP	I-5	...	T→G→ins 59→ter 75
Berkeley, Calif	82	6	2	SP	I-5	...	T→G→ins 59→ter 75
Utah-2	2301	5	2	FS	11	270	*926 ins 11→*ter 301
Tumor-3¶	...	1	0	FS	11	270	*926 ins 11→*ter 301
MSKCC/Utah-1	...	1	0	FS	11	339	*1128 ins A→*ter 345
Montreal, Quebec	270	4	3	FS	11	339	*1128 ins A→*ter 345
Strang Center, New York, NY	2903	1	2	FS	11	361	*1201 del 11→*ter 361
University of Lund, Sweden	8	1	8	FS	11	392	*1294 del 40→*ter 398
Montreal, Quebec	185	1	3	FS	11	392	*1294 del 40→*ter 398
Michigan	32	4	1	NS	11	562	Gln 562 ter
University of Lund, Sweden	9	4	2	FS	11	655	*2080 ins A→*ter 672
USC-1	052	5	1	FS	11	725	*2294 del G→*ter 735
Strang Center, New York, NY	2802	2	2	FS	11	766	*2415 del AG→*ter 766
Berkeley, Calif	3	3	4	FS	11	826	Thr 826 Lys
Montreal, Quebec	179	2	3	MS	11	826	*2477 del C→*ter 845
Stockholm, Sweden	AL48	3	1	FS	11	826	*2477 del C→*ter 845
Stockholm, Sweden	BR33	5	1	FS	11	894	*2800 del A→*ter 901
Berkeley, Calif	1	16	2	FS	11	915	*2863 del TC→*ter 915
Berkeley, Calif	102	5	3	FS	11	956	*2982 del 5→*ter 968
Utah-2	2305	2	7	FS	11	1002	*3121 del A→*ter 1023
Montreal, Quebec	218	5	1	FS	11	1016	*3166 ins 5→*ter 1025
University of Lund, Sweden	24	1	6	FS	11	1016	*3166 ins 5→*ter 1025
Stockholm, Sweden	BR24	2	1	NS	11	1203	Arg 1203 ter
Berkeley, Calif	74	6	0	NS	11	1250	Glu 1250 ter
Michigan	161	3	2	FS	11	1252	*3875 del 4→*ter 1262
Michigan	15	9	2	FS	11	1259	*3896 del T→*ter 1263
ICRF	BC3	7	1	FS	11	1290	*3986 del AA→*ter 1293
CRC Group, Cambridge, England	OV25	2	3	NS	11	1313	Gln 1313 ter
Utah-1	2082	25	21	FS	11	1355	*4184 del 4→*ter 1364
Montreal, Quebec	183	4	1	FS	11	1355	*4184 del 4→*ter 1364
Berkeley, Calif	2	5	0	FS	11	1355	*4184 del 4→*ter 1364
Institute of Cancer Research, Sutton, England	B229	2	1	FS	11	1355	*4184 del 4→*ter 1364
Institute of Cancer Research, Sutton, England	B227	3	1	FS	11	1355	*4184 del 4→*ter 1364
CRC Group, Cambridge, England	OV34	3	4	MS	13	1443	Arg 1443 Gly
Michigan	51	3	0	NS	13	1443	Arg 1443 ter
Strang Center, New York, NY	1900¶	3	1	NS	15	1541	Glu 1541 ter
Tumor-2¶	...	1	0	MS	16	1637	Pro 1637 Leu
Tumor-1¶	...	0	1	FS	16	1656	*5085 del 19→*ter 1671
Strang Center, New York, NY	8622#	4	1	MS	18	1708	Ala 1708 Glu
Tumor-2¶	...	1	0	FS	20	1756	*5382 ins C→*ter 1829
Utah-1	1910	4	0	FS	20	1756	*5382 ins C→*ter 1829
Montreal, Quebec	101	2	2	FS	20	1756	*5382 ins C→*ter 1829
Montreal, Quebec	162	3	1	FS	20	1756	*5382 ins C→*ter 1829
Montreal, Quebec	166	5	2	FS	20	1756	*5382 ins C→*ter 1829
Montreal, Quebec	279	4	0	FS	20	1756	*5382 ins C→*ter 1829
Glasgow-1	E-001**	9	1	FS	20	1756	*5382 ins C→*ter 1829
MSKCC	25-21	0	3	FS	20	1773	*5438 ins C→*ter 1829
Michigan	202	14	6	MS	21	1775	Met 1775 Arg
Utah-1	2099	17	2	MS	21	1775	Met 1775 Arg
Tumor-2¶	MC44	1	0	FS	24	1853	*5677 ins A→*ter 1853
Berkeley, Calif	77	7	0				

*MSKCC indicates Memorial Sloan-Kettering Cancer Center, New York, NY; USC, University of Southern California, Los Angeles; and ICRF, Imperial Cancer Research Fund, London, England. Ellipses indicate sporadic samples without family number.

†Cases with breast-ovarian multiple primary cancers are counted once in each column.

‡IR indicates inferred regulatory; FS, frameshift; MS, missense; SP, splice site; and NS, nonsense.

§For missense and nonsense mutations, the mutation description contains wild-type amino acid, affected codon, altered amino acid (or ter). For frameshift mutations, the format is nucleotide, insertion or deletion, specific nucleotides changed (if <3) or number inserted or deleted (if >2), and the amino acid in which the frameshift results in a termination signal. Nucleotides refer to the *BRCA1* complementary DNA sequence in GENBANK under accession number U-14680.

‖The mutation in this family was independently identified in both the ICRF and Glasgow laboratories.

¶All five mutations identified in tumors were also found in the germline of those individuals.

#The mutation in this family was independently identified by both Myriad Genetics, Salt Lake City, Utah, and the University of Pennsylvania, Philadelphia, laboratories.

**This family is identical to CRC B082 in which the identical mutation was identified by the CRC Group.

Table 4.—Frequency of Two Common *BRCA1* Mutations Detected Through Targeted Screening in Defined Samples

Sample Set*	No. Studied	No. of Mutations Found	
		185 del AG	*5382 ins C*
USC-2	59	4	1
MSKCC/Utah-2	109	3	0
Glasgow-1†	48	1	0
Glasgow-2	98	0	3
Glasgow-3	100	Not tested	2
CRC-ovarian	250	Not tested	1
Milan, Italy	50	1	1

*USC indicates University of Southern California, Los Angeles; MSKCC, Memorial Sloan-Kettering Cancer Center, New York, NY; and CRC, CRC Group, Cambridge, England.

†Excludes one family in which the *5382 ins C* was found by direct sequencing.

Table 5.—Observed Frequency of Different Types of Mutations

Mutation Type	Distinct Mutations, No. (%)* (n=38)	All Mutations, No. (%)† (n=63)
Frameshift	23 (61)	45 (71)
Nonsense	6 (16)	6 (10)
Missense	7 (18)	9 (14)
Other	2 (5)	3 (5)

*Identical mutations are counted only once in this column.

†Each sample in which a mutation has been identified is counted in this column.

intron/exon boundaries and by finding polymorphic sites in genomic DNA reduced to monomorphic sites in cDNA. Two relatively common mutations were found, and their frequencies in other samples were examined by ASO analysis (Table 4). Table 5 describes the distribution of *BRCA1* mutations by type of mutation. The majority of mutations identified were frameshifts that accounted for 58% of the 36 different mutations found to date in *BRCA1* and in 71% of all samples shown to have *BRCA1* mutations through complete screening. The Figure depicts the location of each observed mutation within *BRCA1* and indicates the frequencies of each recurrent mutation. Two tests for a random, uniform distribution of mutations across *BRCA1* were performed, one on all 63 observed mutations and one on the 36 unique mutations observed in the coding sequence. The first test showed a significant departure from a uniform distribution ($P=.02$), reflecting the multiple occurrences of the three frequent mutations. However, the locations of the 36 unique mutations did not depart from a uniform distribution ($P>.20$).

In an attempt to identify a genetic basis for the statistical observation of heterogeneity of ovarian cancer risk described by Easton et al,[16] we examined the distribution across *BRCA1* of the mutations from the 16 families with a high proportion of ovarian cancer and the 31 families with a relatively low proportion of ovarian cancer. Although mutations found in samples from each of the three groups of families defined by the proportion of affected women who had ovarian cancer were found throughout the *BRCA1* coding sequence, mutations in the 3′ portion of the gene may be associated with families with a lower proportion of ovarian cancers in affected women. For example, only four (25%) of the 16 families with a relatively high proportion of ovarian cancer have mutations that lie in the 3′ third of the *BRCA1* gene, while 52% (16/31) of those families denoted as having a low ovarian cancer prevalence have mutations that are found in this region ($P=.08$). To test this hypothesis across the entire *BRCA1* sequence we compared the spatial distributions of the locations of mutations found in the high-proportion families with those from low-proportion families using a two-sample Kolomogorov-Smirnov test. This comparison failed to show a significant difference between the two distributions of mutation locations ($P=.20$).

COMMENT

Mutational Spectra in *BRCA1*

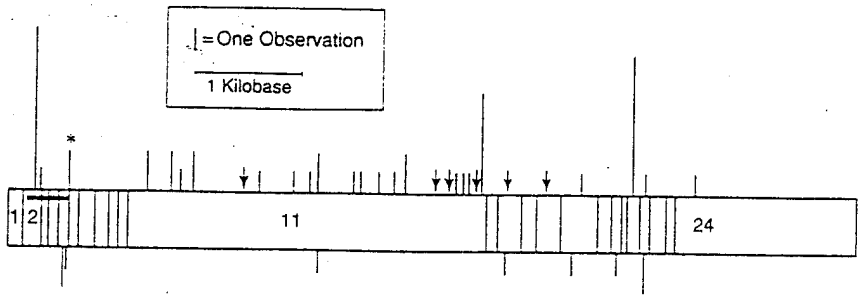
From the data shown in Tables 3 and 4, it is clear that mutations in *BRCA1* have now been found in a diverse set of samples by several different laboratories using a variety of screening methods. Moreover, mutations have been found in many different regions of the gene; phenotypically severe mutations have been found both in the extreme 5′ end of *BRCA1* and in the extreme 3′ portion of the gene. One such mutation found in a family with seven early-onset breast cancer cases produces a protein that is only missing the terminal 10 amino acids, indicating that this region of *BRCA1* plays a role in normal gene function. It is noteworthy that the overwhelming majority of alterations in *BRCA1* have been either frameshift or nonsense mutations. Of the total of 63 independent individuals with mutations identified by complete screening of the coding sequence, 54 (86%) are frameshift, nonsense, splice, or regulatory mutations (Table 5), which result in an truncated protein product. Although the majority of mutations in *BRCA1* identified to date are frameshifts, it may be that these mutations are simply more readily detectable. When exhaustive searches for mutations in samples from a large set of high-risk families that have a very high likelihood of harboring *BRCA1* mutations have been completed, more precise estimates of the frequency of each type of mutation will be obtained.

There is as yet no apparent clustering of mutations in *BRCA1* based on the data depicted in the Figure, using the Kolmogorov-Smirnov test described herein. Twenty-two of the 38 distinct mutations were found in exon 11; however, because this exon contains approximately 60% of the coding sequence of *BRCA1*, this is almost exactly what we would expect, assuming a uniform distribution across *BRCA1*. The only possible evidence of clustering is found in the RING finger motif, which represents 2.3% of the total coding sequence but contains four different mutations (11%).

Frequency of Recurrent Mutations

BRCA1, like many other cancer predisposing genes, is characterized by some mutations that occur relatively frequently. In *BRCA1*, to date, three mutations appear to be relatively common: The *5382 ins C* mutation in codon 1756, the *185 del AG* mutation in codon 23, and the *4184 del 4* mutation in codon 1355 were identified by direct sequencing or SSCP in seven (11%), eight (13%), and five (8%) of the 63 probands in which mutations have been identified, respectively. In addition to these three common mutations, eight additional mutations have been found twice by a complete screen of the cDNA. The total fraction of mutations that are recurrent is 31%, and these 11 recurrent mutations comprise 57% of the 63 mutations detected in probands. Simard et al[18] found that the multiple occurrences of the *5382 ins C* and *185 del AG* mutations shared a common ancestral haplotype for genetic markers within or close to *BRCA1*, indicating that either these mutations are old or that the families are more closely related than originally believed. It will be interesting to perform similar haplotype comparisons among cases of these mutations from the additional families reported herein with these two mutations and the other recurrent mutations.

Many of the probands screened to date for *BRCA1* mutations were selected for having a high prior probability of hav-



BRCA1 transcript showing the locations of mutations shown in Table 3. The *BRCA1* transcript contains 24 exons (shown as boxes), but exons 1 and 4 do not contain coding sequence. The solid black bar extending from exon 2 to exon 5 indicates the location of the RING finger homology. Frameshift mutations occur at locations shown with vertical lines above the complementary DNA diagram, and missense mutations are shown similarly below the diagram. The vertical line topped with an asterisk indicates the splice mutation in intron 6 that inserts 59 base pairs to exon 6 and results in a frameshift. The height of each vertical line corresponds to the number of times each mutation was observed. The arrows above the diagram refer to nonsense mutations, none of which were observed more than once. Thus, all mutations shown above the diagram result in premature termination of the *BRCA1* protein.

ing *BRCA1* mutations. Thus, the mutations found in this set may not be representative of those that would be identified in other sets of patients. However, the two most frequent *BRCA1* mutations have been found multiple times in targeted screening in sets of probands who were, for the most part, either unselected for family history or ascertained with minimal family history (Table 4). At present, nine (2.5%) of 364 samples from breast or ovarian cancer patients screened by ASO have been found to have the *185 del AG* mutation and another eight (1%) of 814 had the *5382 ins C* mutation. It is important to note that six of these mutations were found in premenopausal breast cancer cases selected without regard to family history and that three mutations were found in unselected ovarian cancer cases. Thus, the mutations that were common in the linked and otherwise high-risk families also have been found repeatedly in probands who are more representative of a general clinic population.

Factors Affecting Sensitivity

There are a number of families reported in which the genetic linkage evidence for involvement of *BRCA1* is very strong but for which mutations have not been identified. For example, Michigan families 28 and 30 that have multiple cases of ovarian cancer and convincing evidence for linkage[18] fall into this category, as do a number of families studied by the Berkeley group. Although these could be due to coding mutations missed by SSCP, similar families have been found among groups using direct sequencing (eg, family 102 in Simard et al[16] and kindred 1925 in Miki et al[5]). Such families could be analyzed initially for the presence of a single transcript (indicating that on one chromosome, the copy of *BRCA1* is not being properly transcribed from genomic DNA into messenger RNA) through analysis of the polymorphisms described in the region in cDNA and genomic DNA[5] or by a functional assay when available. The frequency of these presumed regulatory mutations lying outside of the regions currently being analyzed (exons and intron/exon junctions) will affect the utility of any sequence-based test for *BRCA1* mutations.

Factors Affecting Specificity

In the preceding paragraphs, we have focused on factors affecting the sensitivity of an assay for *BRCA1* mutations, ie, the chance that a true, clinically meaningful alteration in the *BRCA1* gene will be detected using a particular screening method. However, we also must consider the specificity of such an assay, ie, the rate of false positives. In practical terms this usually involves the proper interpretation of observed missense mutations. How can one be certain that an observed amino acid substitution found in a patient or relative of a patient is clinically meaningful? Missense mutations, which are recurrent in cancer cases but absent or very rare in the general population, are likely to be clinically meaningful; cosegregation of the missense mutation with the disease in relatives of the proband provides further support that the missense mutation is indeed causal. Unfortunately, there may still be ambiguity regarding the causal nature of missense mutations. This difficulty is exemplified by a *Ser 1040 Asn* putative missense mutation, which was described as a polymorphism by Castilla et al,[19] where it was found not to segregate with the disease and was also detected in three of 232 control chromosomes. In the report of Friedman et al,[20] however, this alteration did segregate with the disease in the family in which it was identified and was absent in 120 control chromosomes. Another example is provided by Utah family K2039 in which a missense mutation changing an arginine to a glycine in amino acid 1347 was detected and found to be absent in 156 control chromosomes. However, this same variant was found in a patient studied by the Berkeley group who also had a severe frameshift mutation; thus, its functional significance is questionable, and it was not included as a mutation in Table 3. A functional assay, if available, could be useful for distinguishing causal mutations from neutral polymorphisms in cases in which the interpretation of a missense mutation is problematic. The current indications that a large fraction of mutations lead to truncation of the *BRCA1* protein may allow use of a truncation assay for *BRCA1* mutations. Such an assay has been described for mutation screening of the *APC* gene.[21] Before the truncation assay is feasible for large-scale screening we need to determine the actual frequency of truncating mutations; the current estimate may be biased due to the ascertainment criteria used in testing probands.

Allelic Heterogeneity

An important question with implications for providing counseling based on *BRCA1* testing is the existence of allelic heterogeneity. That is, do different mutations, or classes of mutations, confer different age-specific risks of breast and/or ovarian cancer. Although these questions can be most definitively addressed through prospective follow-up of relatives of patients with mutations identified through screening of population-based case series, we have attempted to answer one such question based on the data provided in this article. Given the statistical evidence of heterogeneity of ovarian cancer risk from linkage studies of breast/ovarian cancer families, we attempted to identify whether mutations in specific regions of *BRCA1* might be associated with the occurrence of a particularly high or low number of ovarian cancers relative to the number of breast cancers. Based on this analysis, there was not a statistically significant difference in the distribution of the *BRCA1* mutational spectra among low- and high-prevalence ovarian cancer families. However, there was some evidence that mutations nearer the 3′ end of the gene were associated with families with a lower proportion of ovarian cancers in affected women. It is clear that much more data will be necessary to address this question with any degree of preci-

sion. Mutations identified in samples from probands selected with a well-defined mode of ascertainment and with subsequent follow-up of all mutation carriers within the family will be especially useful in this regard.

Interpretation of a Negative *BRCA1* Test Result

It should be emphasized that, at present, failure to find a mutation in an at-risk individual is only useful clinically if a *BRCA1* mutation has been identified in an affected first-degree relative. At-risk women who have not inherited the mutant *BRCA1* allele present in their affected relative would have breast and ovarian cancer risk at rates equal to those in the general population. Families in which no *BRCA1* mutation has been identified will have a risk that combines the probability of nondetection with risks based on empiric data for women with a particular family history, such as those provided by Claus et al.[22] In all cases, risk calculations also should incorporate nongenetic risk factors, particularly if they are shown to interact with the genetic susceptibility conferred by *BRCA1*. A more detailed discussion of risk assessment for breast cancer and associated social and ethical issues have been described[23] and are discussed in an accompanying article in this issue.[6]

In this article, we have deliberately avoided making any specific recommendations as to the most effective method of screening for *BRCA1* mutations, and we did not make recommendations regarding what group of women should be targeted for screening. This is largely because, even with the relatively large number of mutations detailed herein, the variety of methods used and the heterogeneity of the patient samples with regard to family history make any specific determinations of sensitivity and specificity virtually impossible. Rather, we have taken the approach of delineating the relevant issues, discussing the data that are currently available, and discussing how those data may affect those issues. We have included Table 1, however, as a preliminary guide to the a priori risks that a given patient with a particular family and personal history of cancer will harbor a *BRCA1* mutation. Final determination of the relevant issues in a clinically meaningful way must await further data.

This article has, of necessity, focused on diagnostic issues involved in testing exclusively for *BRCA1* mutations. However, a second locus associated with predisposition to early-onset breast cancer, *BRCA2*, has now been localized to chromosome 13q,[24] and there is evidence that a third such locus exists as well. When these genes are isolated, a diagnostic test for breast cancer susceptibility rather than susceptibility due to a specific gene can be offered.

From Myriad Genetics, Salt Lake City, Utah (Drs Shattuck-Eidens, Liu, Fujimura, and Skolnick and Ms McClure); Department of Molecular Endocrinology, CHUL Research Center and Laval University, Quebec City, Quebec (Drs Simard and Labrie); Department of Medical Genetics, McGill University, Montreal, Quebec (Dr Narod); Department of Hematology-Oncology, University of Pennsylvania, Philadelphia (Drs Couch, Hoskins, and Weber); National Center for Human Genome Research, Bethesda, Md (Drs Castilla and Brody and Mr Erdos); Departments of Molecular and Cell Biology and School of Public Health, University of California, Berkeley (Drs Szabo and King and Mss Friedman and Ostermeyer); Departments of Human Genetics (Drs Jhanwar and Offit) and Oncology (Drs Norton and Gilewski), Memorial Sloan-Kettering Cancer Center, New York, NY; Strang Cancer Prevention Center, New York, NY (Drs Lubin and Osborne); Beatson Institute for Cancer Research, Glasgow, Scotland (Drs Black and Ms Boyd); University of St Andrews (Scotland) (Dr Steel); Department of Epidemiology, University of Southern California, Los Angeles (Dr Haile and Ms Ingles); Department of Clinical Genetics, Karolinska Institute, Stockholm, Sweden (Dr Lindblom); Institute of Oncology, University of Lund (Sweden) (Drs Borg and Olsson); Somatic Cell Genetics Laboratories, Imperial Cancer Research Fund, London, England (Drs Bishop and Solomon); Divisions of Experimental Oncology (Dr Radice) and Surgical Gynecological Oncology (Dr Spatti), Instituto Nazionale Tumori, Milan, Italy; CRC Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge (England) (Drs Gayther and Ponder); Section of Molecular Carcinogenesis, Institute of Cancer Research, Sutton, England (Dr Stratton and Mr Warren); and Genetic Epidemiology Group, Department of Medical Informatics, University of Utah School of Medicine, Salt Lake City (Drs Lewis, Skolnick, and Goldgar).

This research was supported in part by National Institute of Health grants CA-55914 to Dr Skolnick and CA-27632 to Dr King and by the Nordic Cancer Union.

We gratefully acknowledge the technical contributions of Steve Bayer, MS, Cheryl Frye, Thanh Tran, Ken Eddington, Jane Feldhaus, Zahra Ghoiami, Bert Bogden, Robert Kehrer, MS, Teresa Janecki, and Robert Olson.

References

1. Kelsey JL, Gammon MD. The epidemiology of breast cancer. *CA Cancer J Clin.* 1991;41:147-165.
2. Newman B, Austin MA, Lee M, King M-C. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high risk families. *Proc Natl Acad Sci U S A.* 1988;85:3044-3048.
3. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the Cancer and Steroid Hormone study. *Am J Hum Genet.* 1991;48:232-241.
4. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990;250:1684-1689.
5. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the 17q-linked breast and ovarian cancer susceptibility gene *BRCA1*. *Science.* 1994;266:66-71.
6. Hoskins KF, Stopfer JE, Calzone KA, et al. Assessment and counseling for women with a family history of breast cancer: a guide for clinicians. *JAMA.* 1995;273:577-585.
7. Lovering R, Hanson IM, Borden KLB, et al. Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc Natl Acad Sci U S A.* 1993;90:2112-2116.
8. Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics.* 1989;5:874-879.
9. Sheffield VC, Fishman GA, Beck JS, Kimura AE, Stone EM. Identification of novel rhodopsin mutations associated with retinitis pigmentosa by GC-clamped denaturing gradient gel electrophoresis. *Am J Hum Genet.* 1991;49:699-706.
10. White MB, Carvalho M, Derse D, O'Brien SJ, Dean M. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics.* 1992;12:301-306.
11. Roest P, Roberts R, Sugino S, van Ommen GJB, den Dunnen JT. Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum Mol Genet.* 1993;2:1719-1721.
12. Grompe M. The rapid detection of unknown mutations in nucleic acids. *Nat Genet.* 1993;5:111-117.
13. Easton DF, Bishop T, Ford D, Crockford GP, and the Breast Cancer Linkage Consortium. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am J Hum Genet.* 1993;52:678-701.
14. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in *BRCA1* mutation carriers. *Lancet.* 1994;343:692-695.
15. Easton DF, Narod SA, Ford D, Steel M. The genetic epidemiology of *BRCA1*. *Lancet.* 1994;344:761.
16. Easton DF, Ford D, Bishop DT. Breast Cancer Linkage Consortium (1995): breast and ovarian cancer incidence in *BRCA1* mutation carriers. *Am J Hum Genet.* 1995;56:265-271.
17. Futreal PA, Liu Q, Shattuck-Eidens D, et al. *BRCA1* mutations in primary breast and ovarian carcinomas. *Science.* 1994;266:120-122.
18. Simard J, Tonin P, Durocher F, et al. Recurrent mutations of the *BRCA1* gene in Canadian breast and ovarian cancer families. *Nat Genet.* 1994;8:392-398.
19. Castilla L, Couch F, Erdos M, et al. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nat Genet.* 1994;8:387-391.
20. Friedman LS, Ostermeyer EA, Szabo CI, et al. Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat Genet.* 1994;8:399-404.
21. Van der Luijt R, Meera Khan P, Vasen H, et al. Rapid detection of translation-terminating mutations at the adenomatous polyposis coli (*APC*) gene by direct protein truncation test. *Genomics.* 1994;20:1-4
22. Claus EB, Risch N, Thompson WD. Age of onset as an indicator of familial risk of breast cancer. *Am J Epidemiol.* 1990;131:961-972.
23. Offit K, Brown K. Quantitative risk counseling for familial cancer: a resource for clinical oncologists. *J Clin Oncol.* 1994;12:1724-1736.
24. Wooster R, Neuhausen S, Mangion J, et al. Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q 12-13. *Science.* 1994;265:2088-2090.

Common origins of *BRCA1* mutations in Canadian breast and ovarian cancer families

J. Simard[1], P. Tonin[2], F. Durocher[1], K. Morgan[2,3], J. Rommens[4], S. Gingras[1], C. Samson[1], J.-F. Leblanc[1], C. Bélanger[1], F. Dion[2], Q. Liu[5], M. Skolnick[5,6], D. Goldgar[6], D. Shattuck-Eidens[5], F. Labrie[1] & S.A. Narod[2]

Women who carry mutations in the *BRCA1* gene on chromosome 17q have an 85% lifetime risk of breast cancer, and a 60% risk of ovarian cancer. We have identified *BRCA1* mutations in 12 of 30 (40%) Canadian families with breast and/or ovarian cancer, including six of the eight families (75%) that contained two cases of early-onset breast cancer and two cases of ovarian cancer. Six frameshift mutations account for all 12 mutant alleles, including nucleotide insertions (two mutations) and deletions (four mutations). Four independent families carried the same 1 basepair (bp) insertion mutation in codon 1755 and four other families shared a 2 bp deletion mutation in codons 22–23. These families were not known to be related, but haplotype analysis suggests that the carriers of each of these mutations have common ancestors.

[1]*Laboratory of Molecular Endocrinology, CHUL Research Center and Laval University, 2705, boul. Laurier, Quebec City, Quebec, Canada, G1V 4G2* [2]*Division of Medical Genetics, Department of Medicine and Department of Human Genetics and* [3]*Department of Epidemiology and Biostatistics, McGill University, 1650 Cedar Avenue, Montreal, Quebec, Canada, H3G 1A4* [4]*Department of Genetics, Hospital for Sick Children and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada* [5]*Myriad Genetics Inc., Salt Lake City, Utah, USA* [6]*Department of Medical Informatics, University of Utah, Salt Lake City, Utah, USA*

Correspondence should be addressed to S.A.N.

In 1990, a gene for hereditary breast and ovarian cancer, designated *BRCA1*, was mapped to chromosome 17q by linkage analysis[1,2]. Earlier this year, it was identified by positional cloning[3]. It was estimated that 92% of families which contain two or more cases of early-onset breast cancer, and two or more cases of ovarian cancer carry mutations in *BRCA1* (ref. 4). Miki and colleagues were able to identify *BRCA1* mutations in five of six extended breast cancer kindreds which showed linkage to the *BRCA1* locus[3]. The *BRCA1* gene contains 22 coding exons, which are distributed over approximately 100 kilobases (kb), and codes for a protein of 1,863 amino acids. The function of the BRCA1 protein is not yet known, but it is predicted to contain a zinc finger domain which is characteristic of proteins which regulate gene transcription.

To establish the frequency and spectrum of mutations in the *BRCA1* gene in families with the breast and ovary cancer syndrome, we have sequenced the *BRCA1* coding region in an index case from 30 breast cancer and breast-ovarian cancer families from Canada. We now describe six novel mutations in *BRCA1* in 12 of these families. Haplotype analysis of intragenic and flanking markers provided evidence for a common origin for each of the mutations that was observed in several families.

Mutations in the *BRCA1* gene and haplotypes

The 30 families with early-onset breast cancer and/or ovarian cancer were identified through the Hereditary Cancer Clinic of McGill University (Table 1). Nine families had positive evidence for linkage to the intragenic *D175855* locus (lod scores greater than 0.25). Six families were uninformative for linkage (lod scores -0.04 to 0.14). Two families gave evidence against linkage (lod scores less than -0.8). The other 13 families were not tested for linkage because DNA was not available from multiple affected individuals.

The exons of the *BRCA1* coding region and the exon–intron splicing boundaries were sequenced in one or more members of each family. For a small number of cases DNA was not available on an affected member. For these, other gene carriers were chosen for sequencing; these were either obligate carriers, or were individuals who had been previously shown to carry the haplotype of risk. To study haplotype association, family members were typed using four microsatellite polymorphisms located within or near *BRCA1*; from centromere to telomere, these are: *D17S855*, *D17S1322*, *D17S1323* and *D17S1327* (refs 6,7) (Table 2).

Direct sequencing of PCR products of the coding region and exon–intron boundaries of *BRCA1* revealed the presence of mutations in 12 of the 30 families studied (Fig 1*a*; Table 3). The mutation designated 185delAG results from a deletion of one of two consecutive AG dinucleotides of the normal sequence TTA GAG of codons 22–23. This mutation changes the reading frame of the mRNA and causes a premature termination codon at position 39 in exon 3. This mutation was detected in index cases from four families (180, 235, 253 and 255) that were not known to be related and which originated from different areas in Canada. In these four families there were a total of 12 cases of breast cancer and 11 cases of ovarian cancer. These four families appear to share a common haplotype (GCCE) for the four markers tested (in families 180 and 253, the haplotypes could not be determined unambiguously because only a single affected individual was typed; however the genotypes were consistent with their being

Table 1 Characteristics of 30 cancer families which have been tested for *BRCA1* mutations

Family	Number of breast cancers	Number of ovarian cancers	Lod score for linkage to *D17S855*	Probability of linkage to *BRCA1* (%)
101	2	2	0.08	95.2
102	5	1	0.68	95.3
113	4	1	–	81
121	2	1	0.01	81.3
130	3	0	–0.04	42.7
136	2	1	–	81
161	4	1	–	81
162	3	1	–	81
164	6	1	–	81
166	5	2	–	81
172	0	1	–	–
178	4	1	0.27	88.8
179	2	3	–	92
180	2	2	–	92
183	4	1	0.53	93.5
185	1	3	0.91	98.9
186	3	0	0.06	–
211	2	1	0.26	88.5
213	3	1	–0.03	79.9
218	5	1	0.33	90.0
235	4	2	–0.82	63.4
240	0	3	0.27	95.5
247	1	3	–	92
248	3	1	–	81
253	1	3	–	92
254	2	2	–1.05	50.6
255	6	7	1.05	99.2
259	2	1	–	81
270	4	3	0.69	98.3
279	4	0	0.14	53.0

Lod scores for linkage between *BRCA1* and the intragenic marker, *D17S855*, were computed using the penetrances as described previously[10]. The probabilities of linkage are estimated by incorporating the lod score and the following prior probability of linkage: 45% for families with no case of ovarian cancer; 81% for families with one case of ovarian cancer; and 92% for families with two or more cases of ovarian cancer.

Table 2 Definitions of alleles with sizes and frequencies

Locus probe	*D17S855* (AFM248yg9)		*D17S1322* (s754)		*D17S1323* (s975)		*D17S1327* (ED2)	
Allele	size	freq. (%)	size	freq. (%)	size	freq. (%)	size	(n)
A	158	–	134	4	161	2	173	
B	156	5	131	3	159	4	171	
C	154	19	128	11	157	19	169	
D	152	18	125	25	155	3	167	(1)
E	150	15	122	51	153	4	165	(2)
F	148	12	119	4	151	68	163	(1)
G	146	26	116	1	149	–	161	(5)
H	144	5			147	–	159	
I	142	–			139	–	157	
J							155	
K							153	
L							135	
M							133	(6)
N							131	
O							129	

Allele sizes are in basepairs. Allele frequencies are derived from ref. 13 (*D17S855*) and from GDB. –, Newly identified alleles. For the *D17S1327* polymorphism, the numbers of observed alleles of particular size were based on 15 control chromosomes. Allele M is the most frequent in the present study and in independnet individuals from CEPH families (0.63).

carriers of the at-risk haplotype). This haplotype was not present in any of the 26 index cases who did not carry this particular mutation (Fisher's exact best, $p<0.0001$).

Four different frameshift mutations were detected in exon 11. The mutation designated 1128insA was detected in the index case of family 270. This mutation consists of an insertion of a single A in the normal sequence GAA AAA AAG of codons 337–339. This mutation leads to a premature termination codon at position 345. The mutation designated 1293del40 was found in family 185 (Fig. 1*b*). It is a 40 bp deletion that includes codons 392–404 and the first nucleotide of codon 405. This mutation leads to a stop codon, 5 codons distal to the deletion and a predicted truncated BRCA1 protein of 396 amino acids in length. In family 218, the mutation designated 3121delA was detected (Fig. 1*b*). This mutation is the result of the deletion of a single A in the normal sequence GAA AAC of codons 1001–1002 and leads to a premature termination codon at position 1023. The fourth exon 11 mutation, designated 4184del4 was found in family 183 (Fig. 1*b*). This is a deletion of the last nucleotide of codon 1355 and all of codon 1356 (normal sequence AAT CAA), and produces a termination codon at position 1364. In the four families with exon 11 mutations, there were 14 cases of breast cancer and eight cases of ovarian cancer.

Four other unrelated families (101, 162, 166 and 279) shared a mutation designated 5382insC in exon 20, consisting of the insertion of a single C in codon 1755. This change results in a stop codon at position 1829 in exon 24. One of these, family 166, was remarkable because ten cases of cancer appeared in a single large sibship, including three cases of breast cancer, two ovarian cancers, two leukaemias, two pancreatic cancers and one prostate cancer (Fig. 2*a*). A case of leukaemia and a case of Hodgkin's disease were seen in more recent generations. Family 279 was also unusual — one obligate carrier was diagnosed with cancer of the breast and of the fallopian tube, and another was diagnosed with cancer of the prostate (Fig. 2*b*). In the four families with the 5382insC mutation, there were 14 cases of breast cancer and five cases of ovarian cancer. The identical

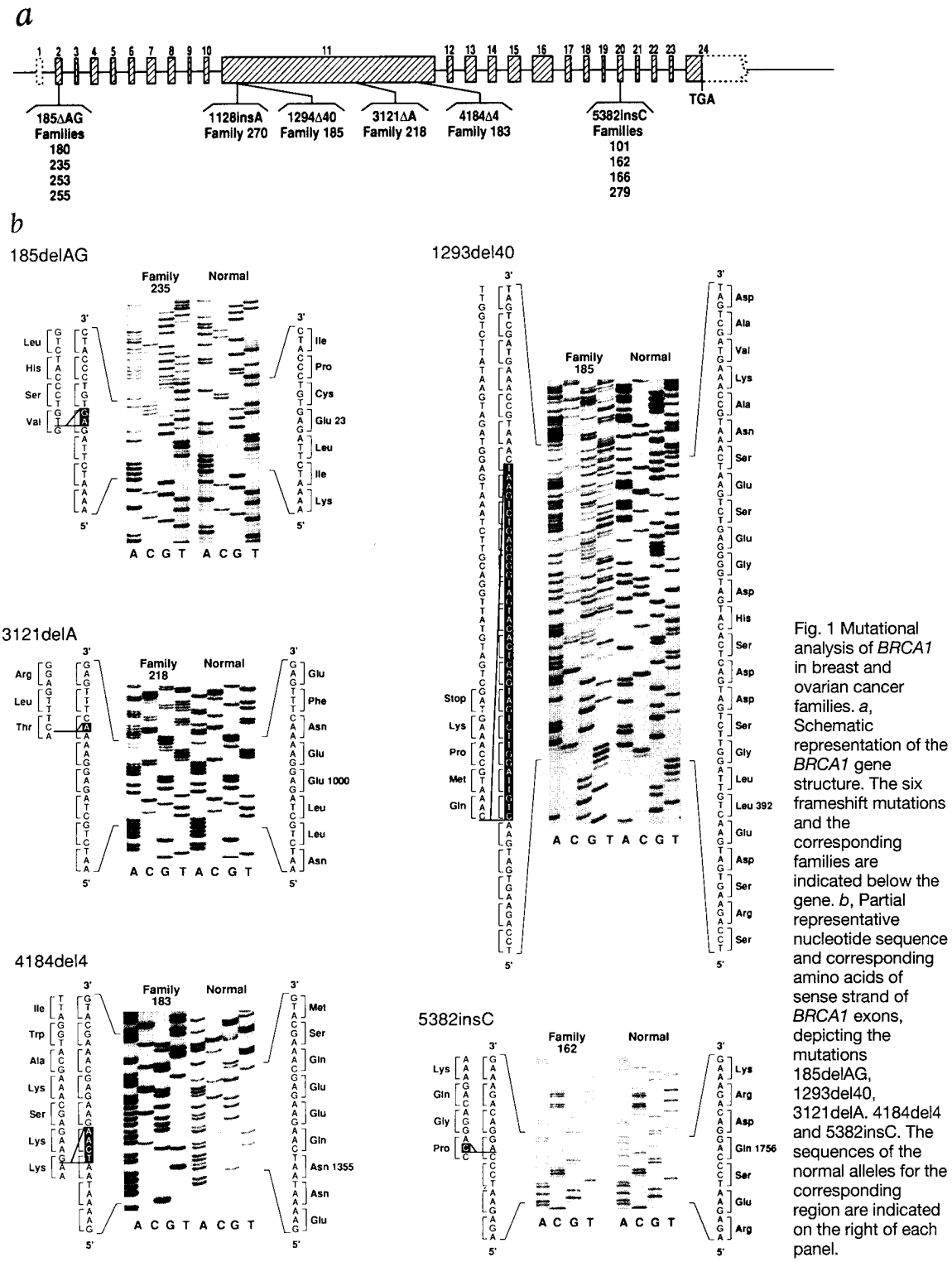


Fig. 1 Mutational analysis of *BRCA1* in breast and ovarian cancer families. *a*, Schematic representation of the *BRCA1* gene structure. The six frameshift mutations and the corresponding families are indicated below the gene. *b*, Partial representative nucleotide sequence and corresponding amino acids of sense strand of *BRCA1* exons, depicting the mutations 185delAG, 1293del40, 3121delA. 4184del4 and 5382insC. The sequences of the normal alleles for the corresponding region are indicated on the right of each panel.

mutation was previously identified in Utah kindred 1910 (ref. 3). The four Canadian families share the haplotype DEFO (the markers in family 166 could not be phased, but the genotype was consistent with carrying the at-risk haplotype). The DEFO haplotype did not appear in any of the 26 index cases who did not carry the 5382insC (Fisher's exact test, $p<0.001$). In fact, the rare O allele of the *D17S1327* polymorphism (129 bp) was not observed on any of the 52 chromosomes from index cases who did not have this particular *BRCA1* mutation nor on 80 chromosomes from independent individuals of CEPH families.

Two families appeared not to be linked to chromosome 17q, but were investigated because each contained two cases of ovarian cancer, and it is has been previously estimated that 92% of families of this type will carry *BRCA1* mutations[4]. In one of these (family 235), a *BRCA1* mutation was found in a woman with ovarian cancer (Fig. 2*c*). Her sister was diagnosed with breast cancer at age 47 and does not carry the 185delAG mutation.

Discussion

We describe six different mutations in the *BRCA1* gene in a total of 12 breast-ovarian cancer families out of 30 families studied (40%), confirming that this gene is responsible for the hereditary breast and ovarian cancer syndrome which has been mapped to chromosome 17q. Of the six mutations, two were each seen four times and four were unique. One of the mutations, 5382insC, found in four Canadian families, was the first identified in a Utah kindred[3]. These authors also identified a deletion mutation, 189del11, in a Utah kindred. We observed a nearby deletion mutation, 185delAG, in four unrelated patients. Both of these mutations in exon 2 are predicted to cause truncation at the beginning of the zinc finger domain of the putative polypeptide.

In general, the families in which mutations were detected were clear examples of the breast ovarian-cancer syndrome. Eight of the 12 families with mutations contained multiple cases of ovarian cancer, in contrast to only four of the 18 families in which mutations were not found (Fisher's exact test p=0.024). Therefore, it was possible to detect a mutation in eight of the 12 families (67%) with two or more cases of ovarian cancer, and in six of the eight families (75%) with two cases of early-onset breast cancer and two cases of ovarian cancer. The average age of breast cancer in the families with a detectable mutation was 43.8 years, versus 55.5 years for the 18 families in which we did not find mutations (Fig. 3; Wilcoxon test of the difference in the age distributions, $p<0.0001$).

We were able to identify mutations in five of nine families with lod scores greater than 0.25 to *D17S855*. It was not possible to identify linkage in all families with mutations; of the 12 families with mutations, linkage analysis was not feasible in four because we were unable to obtain the necessary samples, two other families were uninformative (lods of 0.08 and 0.14) and one family gave a negative lod score (lod=–0.82). For five of the 12 families with mutations, we had positive evidence for linkage (lods ranging from 0.33–1.05). Currently, linkage analysis and direct mutation analysis are complementary, techniques that should be available where genetic assessment is offered to high-risk families.

The nature of the mutations in our other families is not yet known. It is possible that there are *BRCA1* mutations

Table 3 *BRCA1* mutations and corresponding haplotypes

Family	Mutation	Haplotype			
		S855	S1322	S1323	S1327
101	**5382insC**	**D**	**E**	**F**	**O**
102	–	C	E	F	M
113	–	B/E	D/E	F	M
130	–	C	E	F	M
121	–	F/H	D	C	G/H
136	–	F/G	C/D	F/G	B/I
161	–	F	C/E	D/F	B/M
162	**5382insC**	**D**	**E**	**F**	**O/M**
164	–	D/G	B/F	C/E	E/G
166	**5382insC**	**D/G**	**C/E**	**C/F**	**F/O**
172	–	C/D	D/E	E/F	M
178	–	E	C	F	J
179	–	F/G	D/E	B/F	G/M
180	**185delAG**	**G/H**	**C/E**	**C/F**	**E/L**
183	**4184del4**	**A**	**F**	**F**	**M**
185	**1293del40**	**F**	**D**	**F**	**M**
186	–	G	D	C	C
211	–	E	C	C	E
213	–	D	E	E	M
218	**3121delA**	**E**	**E**	**F**	**M**
235	**185delAG**	**G**	**C**	**C**	**E**
240	–	E	E	F	M
247	–	F	D/E	F	M
248	–	A/G	D	C/F	D/M
253	**185delAG**	**G/H**	**C/D**	**B/C**	**E/F**
254	–	D/H	D	C/F	M/C
255	**185delAG**	**G**	**C**	**C**	**E**
259	–	C/F	E/G	E/F	M
270	**1128insA**	**G**	**D**	**B**	**D/I**
279	**5382insC**	**D**	**E**	**F**	**O**

–, No mutation identified. Marker alleles correspond to those in Table 2. Both alleles are given for markers which could not be phased with respect to *BRCA1*.

outside of the coding region which may result in an unstable transcript[3]. It is also possible that some of the family clusters are not due to mutations in *BRCA1*. There are now clear examples of breast and ovarian cancer families which are unlinked to *BRCA1*; a small number of these show linkage to *BRCA2* on chromosome 13 (ref. 8). Another possibility is that we indavertently chose to sequence a sporadic case in a *BRCA1*-linked family.

If BRCA1 protein acts as a tumour suppressor, we would expect a large proportion of the mutations to result in the inactivation of one allele through premature stop codons. In our families, all of the 12 mutations result in frameshifts that lead to premature stop codons, and as predicted from the tumour suppressor model, the mutations are distributed throughout the coding region. Surprisingly, somatic mutations in *BRCA1* have not yet been found in tumour tissue from women with hereditary or non-familial cancers[9].

There were two recurring mutations among the six identified, which together account for two-thirds of the observed mutant alleles. One of these, 5382insC, was strongly associated with a rare allele of the *D17S1327* polymorphism. A wide spectrum of cancers was observed in families with this mutation, including cancer of the fallopian tube, pancreas, prostate and leukaemia and Hodgkin's disease. It will be of interest to see if these observations are confirmed in other families with this mutation.

Our data suggest that screening for *BRCA1* mutations in high-risk women may be facilitated if common mutations are sought first, and if information on linkage disequilibrium is incorporated. A relatively small number of mutant alleles may account for the majority of breast and ovarian cancer families in Canada.

Note added in proof: A missense mutation was identified in family 179. The mutation is a Thr to Lys substitution at codon 826 (ACA→AAA). It was not detected in 86 control individuals

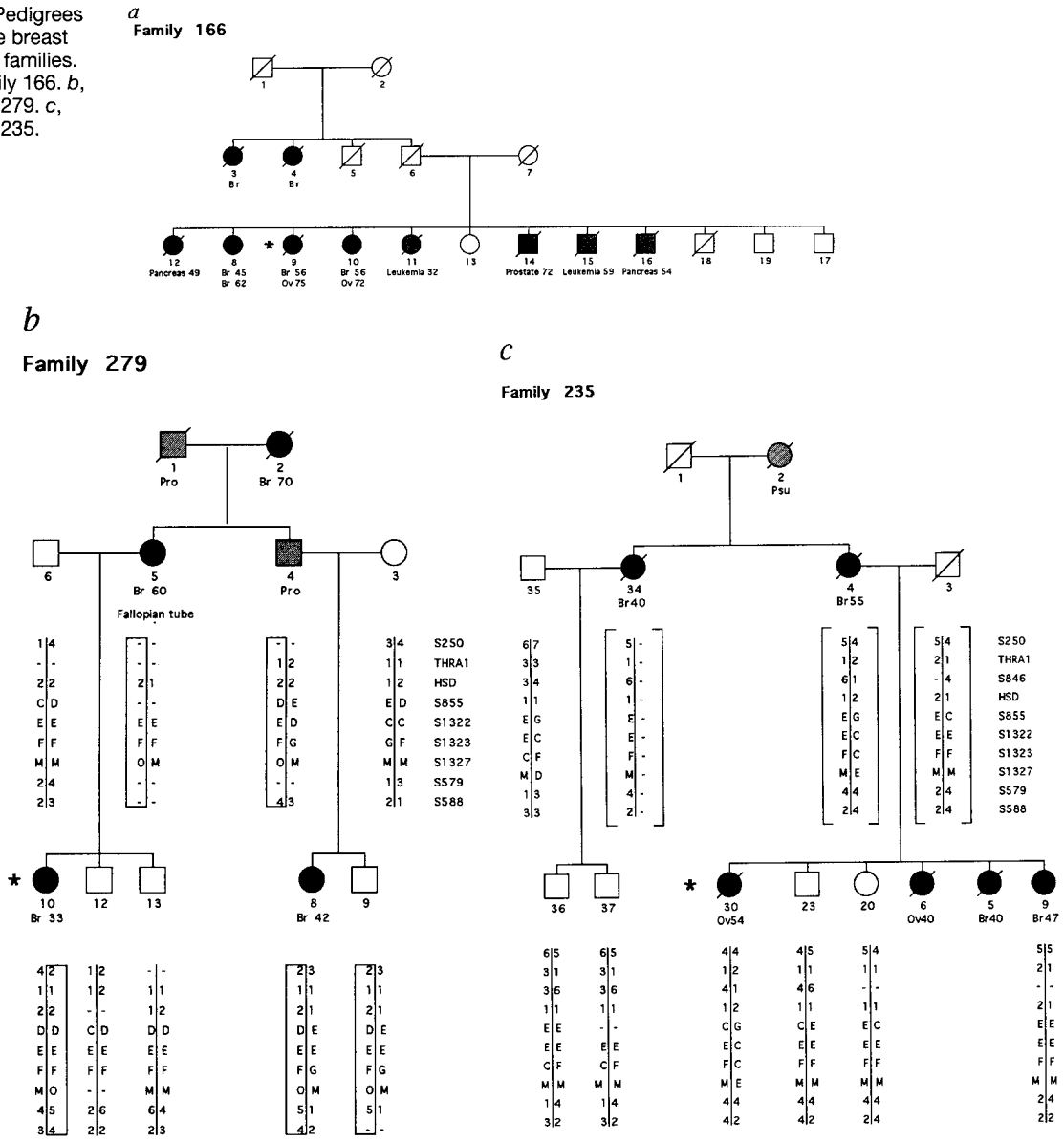


Fig. 2 Pedigrees of three breast cancer families. *a*, Family 166. *b*, Family 279. *c*, Family 235.

Table 4 Oligonucleotide primers for *BRCA1*

Exon #	5' primer	3' primer
2	GAAGTTGTCATTTTATAAACCTTT	TGTCTTTTCTTCCCTAGTATGT
3	CTCAGTTCCTGACACAGCAGAC	GGAGTTGGATTTTCGTTCTCACTTA
4	GTCAAAGAGATAGAATGTGAGC	CCCGTCTCTACAGAAAACAC
5	GGCTCTTAAGGGCAGTTGTGAG	CTTTTCCTACTGTGGTTGCTTCC
6	AAGGTTGATAATCACTTGCTG	GTAATGTGCAAACTTCCTG
7	GCATACATAGGGTTTCTCTTGG	TCGGGTTCACTCTGTAGAAG
8	CATGTTAGCTGACTGATGATGG	GGAATCCAGCAATTATTATTAAATAC
9	TACCTGCCACAGTAGATGCTCAG	AATAGGAAAATACCAGCTTCATAGA
10	TAAGATTGGTCAGCTTTCTGTAATC	ACAGTACTGTATCTACCCACTCTC
11a	CTAAGTGTTCAAATACCAGTGAAC	ACTGGAGCCCACTTCATTAGTAC
11b	GATTTCCACCTCCAAGGTGTATGA	TGTGGCTCAGAAACAAATGCTCC
11c	CAGAGTTCACTCCAAATCAGTAGAGAG	CTGAGTGCCATAATCAGTACCAGG
11d	CCAAGTACAGTGAGCACAATTA	AAATGTGCTCCCAAAAGCATAAA
12	CCAGTCCTGCCAATGAGAAGAAA	TGTCAGCAAACCTAAGAATGT
13	CATTTAATGGAAAGCTTCTCAAAG	AAATGTTGGAGCTAGGTCCTTAC
14	ATTCTAACCTGAATTATCACTATCA	AAAGTGTATAAATGCCTGTATGCAAA
15	TGGCTGCCCAGGAAGTATG	AACCAGAATATCTTTATGTAGGA
16	AATTCTTAACAGAGACCAGAAC	AAAACTCTTTCCAGAATGTTGT
17	GTGTAGAACGTGCAGGATTG	TCGCCTCATGTGGTTTTA
18	GAGGCTCTTTAGCTTCTTAGG	AAAGAGACCCATTTTCCCAGCA
19	CTGTCATTCTTCCTGTGCTC	CATTGTTAAGGAAAGTGGTGC
20	ATATGACGTGTCTGCTCCAC	GGGAATCCAAATTACACAGC
21	AAGCTCTTCCTTTTTGAAAGTC	GTAGAGAAATAGAATAGCCTCT
22	TCCCATTGAGAGGTCTTGCT	GAGAAGACTTCTGAGGCTAC
23	CAGAGCAAGACCCTGTCTC	ACTGTGCTACTCAAGCACCA
24	ATGAATTGACACTAATCTCTGC	GTAGCCAGGACAGTAGAAGGA

Methodology

Families. The 30 families analysed in this study were referred to the McGill University Hereditary Cancer Clinic at the Montreal General Hospital for genetic counselling and for risk assessment. All 30 families are Caucasian and principally reside in Canada. Sixteen families were previously studied for linkage to *BRCA1* (ref. 5). In that study, it was estimated that 94% of families with early-onset breast cancer and ovarian cancer were linked families. All but three families contained cases of ovarian cancer, and all but two contained cases of early-onset breast cancer. Three families were included because they contained a case of primary cancer of the peritoneum (families 130 and 172) or of the fallopian tube (family 279). These cancer sites bear a histologic relationship to epithelial cancers of the ovary. Family 186 was studied because one of the women had multiple primary cancer of the breast and endometrium, and the 17q haplotype was consistent with linkage to *BRCA1* (data not shown).

Linkage analysis. In order to establish probabilites of linkage, lod scores of linkage of predisposition to breast cancer and/or ovarian cancer to *D17S855* were computed using a previously described model[10]. The distance between *D17S855* and *BRCA1* was assumed to be 0 cM (ref. 3). The prior probabilities of linkage used in the calculation were 45%, 81% and 92% for families with 0, 1 and 2 cases of ovarian cancer, respectively[4,11].

Mutation analysis. Genomic DNA was isolated from blood samples. DNA from a representative member of each family was amplified by PCR using 25 pairs of intron-based primers. In most cases, the individual selected was a woman who had been affected with early-onset breast cancer or with ovarian cancer. In some families for which linkage had been established, genomic sequencing was performed on DNA from an unaffected individual who had been shown to carry the *BRCA1* haplotype of risk. Ten families (101, 102, 113, 121, 136, 161, 166, 178, 179 and 270) were analysed by the Utah group by DNA sequencing as previously described[3]. The other 20 families were analysed in Quebec City. PCR was performed using the 26 pairs of primers presented in Table 4, in a 50 µl volume containing 10 mM Tris-HCL, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 50 µM of dNTPs, 0.5 µM of each primer, 0.1% Triton X100, 2% DMSO and 100 ng of genomic DNA. The reaction was first heated at 96 °C for 5 min during which time 1 U of *Taq* polymerase (Perkin-Elmer) was added. The reactions were carried out using a Perkin-Elmer Cetus thermal cycler with a 2-step temperature cycle consisting of 40 s of denaturation at 94 °C and 30 s of annealing at 55 °C; the elongation step was considered to be the time the thermal cycler reached 94 °C from the annealing step. After 35 cycles, a final extension at 72 °C for 2 min was performed. The primers were removed by selective precipitation[12] and the PCR products were then subjected to a 35-cycle asymmetric amplification under the same conditions, except for the use of 10 nM limiting primer. The asymmetric PCR products were then purified by selective precipitation.

Single-stranded DNA produced by asymmetric PCR was sequenced by the dideoxy method using the PCR primer or sequence-specific primers with the T7 sequencing kit (Pharmacia LKB Biotechnologies) as described[12]. Briefly, the single-stranded DNA and the primers were mixed together in the annealing buffer and denatured at 95 °C for 5 min. The samples were then quickly frozen on dry ice-ethanol and the sequencing reactions were performed following the protocol provided by the manufacturer using ^{35}S-dATP or ^{35}S-dCTP as the labelled nucleotide. The *BRCA1* coding region was sequenced entirely for each family except for

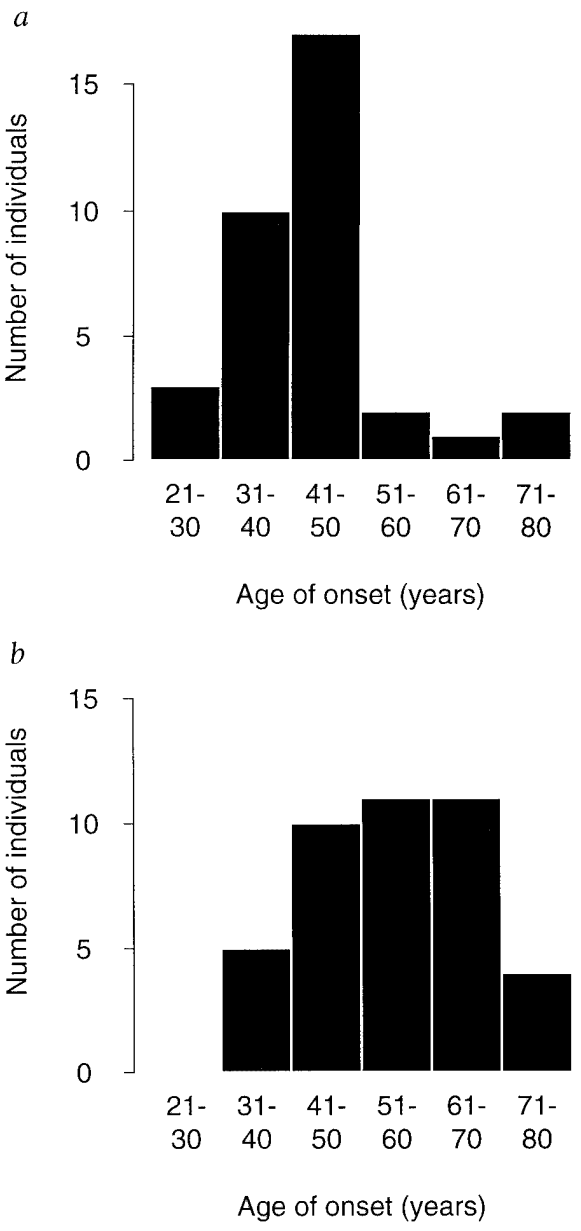


Fig. 3 Distribution of ages of onset of breast cancer in families with and without detectable *BRCA1* mutations. *a*, Individuals from families with a *BRCA1* mutation (n=35). *b*, Individuals from families without a *BRCA1* mutation (n=41).

exons 6 and 7 (89 bp and 140 bp respectively), which were partially sequenced, and for exon 4 which was previously found in only one cDNA isolated from a placental cDNA library[3].

The *BRCA1* mutations were designated according to a suggested nomenclature[14]. The nucleotides were numbered according to the sequence of GenBank entry U14680.

Haplotypes. Haplotypes were constructed for the chromosome 17q region containing the four microsatellites *D17S855* (ref. 13), *D17S1322, D17S1323* and *D17S1327* (refs 6,7) by inspection of segregation patterns, and assuming a minimum number of crossovers. No crossover events were observed in the region spanned by the four microsatellite markers in any of the families studied. The significance of allelic association was assessed by Fisher's exact test.

Received 3 November; accepted 8 November 1994.

1. Hall, J.M. *et al.* Linkage analysis of early onset familial breast cancer to chromosome 17q21. *Science* **250**,1684–1689 (1990).
2. Narod, S.A. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12–23. *Lancet* **338**, 82–83 (1991).
3. Miki, Y. *et al.* A strong candidate gene for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
4. Narod, S.A. *et al.* A heterogeneity analysis of 145 breast-ovarian cancer families. *Am. J. hum. Genet.* (in the press).
5. Tonin, P. *et al.* Linkage analysis of 26 Canadian breast and breast-ovarian cancer families. *Hum. Genet.* (in the press).
6. Goldgar, D.E. *et al.* A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic and genealogic analysis. *J. natn. Cancer Inst.* **86**, 200–209 (1994).
7. Neuhausen, S.L. *et al.* A P1-based physical map of the BRCA1 region from D17S766 to D17S78. *Hum. molec. Genet.* **3**, 1919–1926 (1994).
8. Wooster, R. *et al.* Localization of a breast cancer susceptibility, BRCA2, to chromosome 13q12–13. *Science* **265**, 2088–2090 (1994).
9. Futreal, P.A. *et al. BRCA1* mutations in primary breast and ovarian carcinomas. *Science* **266**, 120–122 (1994).
10. Feunteun, J. *et al.* A breast cancer susceptibility gene maps to chromosome 17q21. *Am. J. hum. Genet.* **52**, 736–743 (1993).
11. Easton, D.F. *et al.* Genetic linkage analysis in familial breast and ovarian cancer — results from 214 families. *Am. J. hum. Genet.* **52**, 678–701 (1993).
12. Simard, J. *et. al.* Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12–21; exclusion of candidate genes EDH17B2 and RARA. *Hum. molec. Genet.* **2**, 1193–1199 (1993).
13. Anderson L. *et al.* High density genetic map of the BRCA1 region of chromosome 17q12–q21. *Genomics* **17**, 618–623 (1993).
14. Beaudet, A.L. & Tsui, L.-C. A suggested nomenclature for designating mutations. *Hum. Mutation* **2**, 245–248 (1993).

Acknowledgements

We thank the family members and the physicians who contributed to this study including R. Margolese, D. Cole, D. Gilchrist, B. McGillivray, B. Rosen and N. Boyd; and R. Moslehi, M. Fujiwara, A.J. Paradis, D. Frappier, G. Leblanc, J.P. Gagné, G. Legris and M. Tranchant for their expert assistance and scientific advice. This work was supported by grants from the Canadian Genetics Diseases Network, the National Cancer Institute of Canada Breast Cancer Research Initiative, Endorecherche and the Department of the Army (US) Grant no. DAMD17-94-J-4299. S.N. is a fellow of the Fonds de la Recherche en Santé du Québec and P.T. is supported by the Cedars Cancer Institute of the Royal Victoria Hospital. J.S. and J.R. are scholars of the Medical Research Council of Canada (MRC). F.D. and S.G. are recipients of studentships from the MRC.

Human Molecular Genetics, 1994, Vol. 3, No. 9 1679–1682

The gene for hereditary breast–ovarian cancer, BRCA1, maps distal to EDH17B2 in chromosome region 17q12–q21

Patricia Tonin, Olga Serova[1], Jacques Simard[2], Gilbert Lenoir[1], Jean Feunteun[3], Kenneth Morgan, Henry Lynch[4] and Steven Narod*

Departments of Medicine and Human Genetics, Montreal General Hospital and McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada, [1]International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon, France, [2]Medical Research Group in Molecular Endocrinology, CHUL Research Centre and Laval University, St Foy, Quebec, Canada, [3]Institut Gustave Roussy, Villejuif, France and [4]Creighton University School of Medicine, Department of Preventive Medicine and Public Health, Omaha, NE, USA

Received June 23, 1994; Revised and Accepted July 19, 1994

A gene for hereditary breast and ovarian cancer, BRCA1, has been mapped to chromosome 17q12–q21. This gene is responsible for cancer susceptibility in the majority of families with multiple cases of ovarian cancer and early-onset breast cancer. We report linkage results of a family with 10 cases of breast cancer and a single case of ovarian cancer. A recombinant event in this family places BRCA1 distal (telomeric) to the locus EDH17B2, which codes for the enzyme estradiol 17β-dehydrogenase II. This recombinant is based on the appearance of breast cancer in a 45 year old woman. Under our genetic model, we estimate the probability that this woman carries a BRCA1 mutation to be 94%. These data further reduce the region of assignment of BRCA1 on chromosome 17q12–q21 and should expedite positional cloning of this important gene.

INTRODUCTION

Susceptibility to hereditary breast and ovarian cancer has been mapped by genetic linkage to a locus, designated BRCA1, on chromosome 17q12–q21 (1,2). BRCA1 has been assigned to the 13 cM interval between the loci D17S250 and D17S588 by a consortium analysis of 214 cancer families (3). Subsequently identified genetic recombinants have narrowed the centromeric boundary of BRCA1 successively to THRA1 (4), to RARA (5), to D17S857 (6), to D17S776 (7) and to D17S702 (8) (the relative position of D17S776 and D17S702 is not known). The telomeric boundary is currently defined at D17S78 (5). The distance between D17S776 and D17S78 is estimated to be 1–2 cM.

This chromosome 17 interval contains the candidate gene EDH17B2, which codes for 17β hydroxysteroid dehydrogenase (17β HSD II). 17β HSD II is responsible for the reversible conversion of estrone, a weak estrogen, into the more potent 17β estradiol and has been considered to be a good candidate gene for BRCA1. However, we and others have sequenced DNA from several unrelated BRCA1 carriers and have failed to show mutations in this gene which would implicate it as the susceptibility gene of the cancer syndrome (5,6). EDH17B1 has a similar genomic structure to EDH17B2, but data suggest that EDH17B1 is a pseudogene (9). To date, no genetic recombinants have been reported between the loci BRCA1 and EDH17B2.

RESULTS

We report here haplotype analysis on breast–ovarian cancer Family 2775 from the Creighton University Family Cancer Registry. This analysis appears to exclude EDH17B2 from the chromosomal region of assignment of BRCA1. The three-point lod score for this family was 1.55 using the closely-linked markers D17S855 and D17S859. This lod score was calculated at D17S855, assuming a 2 cM (sex-averaged) distance between D17S855 and D17S859, under our previously described linkage model (10). Based on a consortium analysis of 48 families, it has been estimated that 79% of families with multiple cases of early-onset breast cancer and a single case of ovarian cancer are linked to BRCA1 (D.Ford, personal communication). Given the positive lod score, the probability that family 2775 is linked to BRCA1 is therefore estimated to be above 99%. A common chromosome 17q haplotype from D17S250 to D17S588 is shared by seven of the eight women with breast cancer in this pedigree who have been typed with these markers (Fig. 1). Individual 128 was diagnosed with invasive breast cancer at age 45. She inherited the marker allele for the D17S855 polymorphism associated with high cancer risk from her father, as well as several other telomeric markers associated with increased risk. However, she inherited allele 2 of the HSD.E polymorphism, and allele 1 of the HSD.del polymorphism from her father. Elsewhere in the pedigree HSD.E allele 1 and HSD.del allele 2 are coupled with the BRCA1 mutation. It appears, therefore, that either individual 87 or individual 126 have inherited a recombinant chromosome 17 (presumably from their affected mother). Individual 128 shares the chromosome 17 markers distal to, but not including, the EDH17B2 locus, with the affected descendants of individual 87. If we infer from her early-onset breast cancer that she inherited the BRCA1 mutation as well, then this recombinant places BRCA1 telomeric to EDH17B2. Further analyses exclude the centromeric markers D17S857 and D17S846 as well. The telomeric markers D17S855, D17S579, D17S509, D17S293 were also informative but not excluded. Two other women in the family (individuals 143 and 12) have also inherited the

*To whom correspondence should be addressed

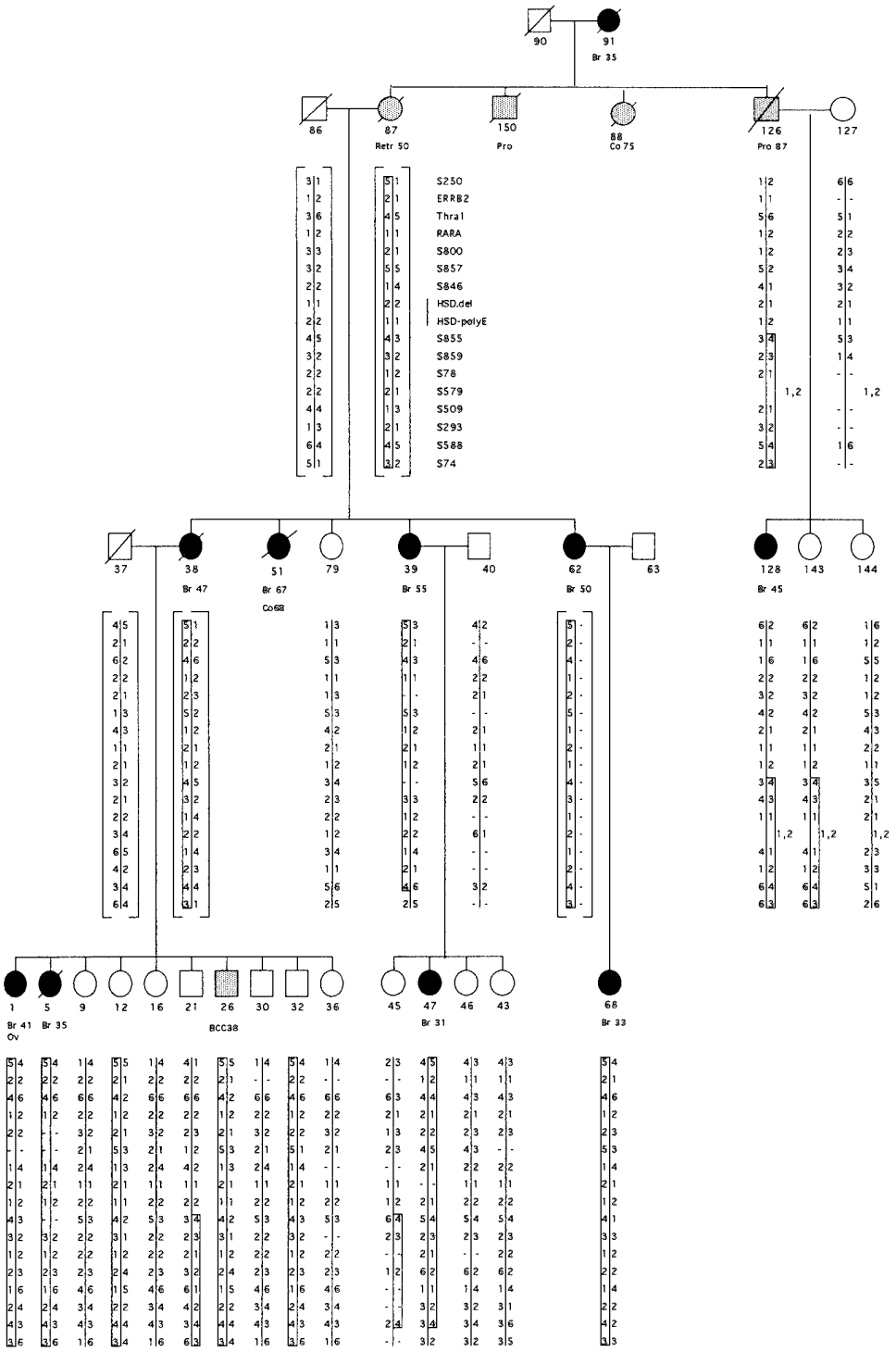


Figure 1. Pedigree of family 2775. Black circles indicate affected women. Diagonal slash indicates deceased. Individual identification numbers appear directly below the symbols. Br = breast cancer, Ov = ovarian cancer, Retr = retroperitoneal cancer, Pro = prostate cancer, Co = colon cancer, BCC = basal cell carcinoma. The numbers following these abbreviations indicate ages of diagnosis. The numbers arranged vertically below the individual symbols indicate the marker alleles arranged into haplotypes. Haplotypes in brackets are inferred. Marker alleles separated by a comma cannot be phased. A dash in the place of a marker typing indicates missing information. Vertical rectangles are used to indicate the haplotype found to be associated with breast or ovarian cancer in the family.

haplotype of risk and have not developed breast or ovarian cancer. Individual 143 underwent a bilateral oophorectomy at age 48 (her ovaries were described as normal). She is currently 52 years old. Individual 12 is currently 31 years old.

DISCUSSION

The recombinant in Family 2775 suggests that BRCA1 is located telomeric to EDH17B2 on chromosome 17q12−q21. This recombinant significantly narrows the minimal region of

Table 1. Polymorphisms used in this analysis

Locus	Marker	Type of Polymorphism	Reference
D17S250	MFD15	CA	13
ERBB2	ERBB2	RFLP (*Mbo*I)	14
THRA1	THRA1	CA	4
D17S800	200ZF4	CA	15
D17S857	OF1	CA	16
D17S846	UM8	CA	17
EDH17B1	HSD.del	12 bp deletion	18
EDH17B2	HSD.E	RFLP (*Bfa*I)	19
D17S855	248Yg9	CA	15
D17S859	OF4	CA	16
D17S78	p131	RFLP (*Msp*I)	20
D17S579	Mfd188	CA	14
D17S509	clb17.1	RFLP (*Hae*III)	10
D17S293	6C1	CA	14
D17S588	42D6	CA	3
D17S74	CMM86	VNTR (*Hin*fI)	1

assignment of BRCA1 and should expedite efforts to clone this gene. However, because of the possibility of sporadic cases of breast and ovarian cancers, no single recombinant can be taken as proof of BRCA1 placement. Errors in the assignment of the location of BRCA1 have been made in the past based on single cases of breast cancer (e.g. ref. 10). In contrast to our results, Smith and colleagues map BRCA1 to a position centromeric to EDH17B2 (8). However, there was very little evidence that the family of Smith *et al.* was linked to BRCA1, and more recent data reveal that the case of ovarian cancer in this family (who was not typed at the time of the published report) did not share a common chromosome 17q haplotype with the three cases of early-onset breast cancer. Either she is a sporadic case of ovarian cancer or the family is unlinked to BRCA1. A recent report of loss of heterozygosity (LOH), comparing DNA from 130 sporadic breast cancers with constitutional DNA, also suggests a placement of BRCA1 which is inconsistent with our results (11). These investigators identified a minimum common region of loss which was proximal to D17S776, which in turn is reported to be centromeric to EDHB2. The relevance of such LOH studies to mapping BRCA1 is currently unclear.

The evidence that BRCA1 is telomeric to EDH17B2 in Family 2775 is the strongest so far. Linkage to BRCA1 is not in doubt; we estimate that the probability that the family is linked to BRCA1 is above 99%. A more important consideration is that individual 128 may be a sporadic case. Our linkage model (described in reference 10) assigns a relative risk of 60 for the incidence rate of breast cancer in BRCA1 carriers versus non-carriers at the age of 45. Based on pedigree data alone (i.e., not the DNA markers) this model predicts that individual 128 has a 94% chance of having inherited a BRCA1 mutation, versus a 6% chance of being a sporadic case. Other prior probabilities of carrying BRCA1 mutations in this family are estimated to be 44% for individual 143, 93% for individual 126, and 49% for individual 12. Interestingly, prostate cancer, which was diagnosed in individual 126, appears to be increased in frequency among males with BRCA1 mutations (12). Because of the uncertainty of exclusion attributable to a single recombinant, it is important that the present observations be replicated.

We are able to support an assignment of BRCA1 telomeric to the more proximal marker D17S857 in a second breast–ovary

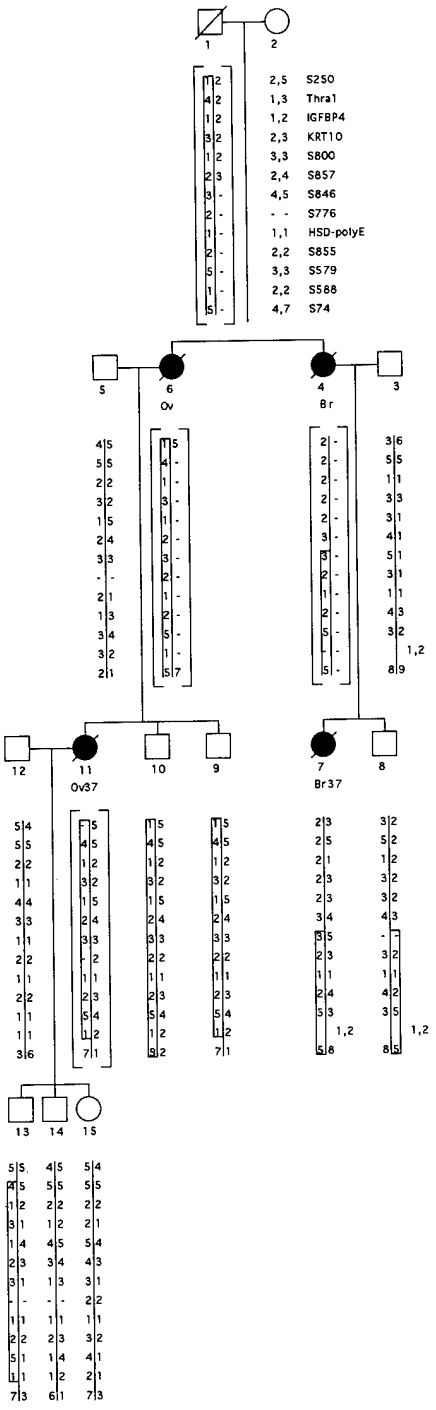


Figure 2. Pedigree of family 101. Symbols as in figure 1.

cancer family, Family 101 (Fig. 2). This family, which contains two cases of early onset breast cancer and two cases of ovarian cancer has a 96% probability of being linked to BRCA1. Including the present observations there have now been six families reported in which recombinants which place BRCA1 distal to D17S857 (6–8).

MATERIAL AND METHODS

Family 2775 was identified in the mid-western United States and is followed at the Creighton University School of Medicine. The family was initially identified for linkage studies because of ten confirmed cases of breast cancer which appeared at an average age of 43.7 years (range 31 to 66 years). Recently, a woman who was previously diagnosed with breast cancer at the age of 41 was found to have ovarian cancer. The family therefore qualifies as a breast–ovarian cancer family. Other cancers in this family include four cases of colon cancer, two cases of prostate cancer and a basal cell carcinoma. Family 101 was identified at McGill University in 1990.

Family members were typed with the polymorphic markers listed in Table 1. The polymorphism HSD.del is located in the 5′ flanking sequence of EDH17B1 and is defined by two alleles which differ in size by 12 base pairs (18). The HSD.E polymorphism is a PCR-based RFLP located in the intron between exons I and II of EDH17B2 (19). Polymorphic markers for the loci IGFBP4, keratin 10, D17S776, SCG43, D17S507 and GIP were also tested in family 2775, but were not informative for the critical recombinant. LOD scores were calculated with the MLINK program (21). The calculation of the posterior probability of linkage is based on a method described previously (22).

ACKNOWLEDGEMENTS

We would like to thank Anne Vivier, France Dion, Marie-France Lavoué, Susan Slominski, Patrice Watson, Theresa Conway and Jane Lynch for helping with the families and for technical expertise. This work was supported by the National Cancer Institute of Canada, the Fonds de la Recherche en Santé du Québec, the Comité Departmental de l'Ain de la Ligue Nationale Française contre le Cancer, the French Ministry of Science and Technology, the Canadian Genetic Diseases Network, Endorécherche, and the Cedars Cancer Institute of the Royal Victoria Hospital.

REFERENCES

1. Hall J.M., Lee M.K., Newman B., Morrow J., Anderson L., Huey B., King M-C. (1990) *Science* **250**, 1684–1689.
2. Narod S.A., Feunteun J., Lynch H., Watson P., Conway T., Lynch J., Lenoir G.M., (1991) *Lancet* **338**, 82–83.
3. Easton D.F., Bishop D.T., Ford D., Crockford G.P. and the Breast Cancer Linkage Consortium. (1993) *Am. J. Hum. Genet.* **52**, 678–701.
4. Bowcock A.M., Anderson L.M., Fridedman L.S., Black D.M., Osborne-Lawrence S., Rowell S.E., Hall J.M., Solomon E., King M-C. (1993) *Am. J. Hum. Genet.* **52**, 718–722.
5. Simard J,. Feunteun J., Lenoir G.M., Tonin P., Normand T., Luu The V., Vivier A., Lasko D., Morgan K., Rouleau G., Lynch H., Labrie F., Narod S.A. (1993) *Hum. Mol. Genet.* **2**, 1193–1199.
6. Kelsell D.P., Black D.M., Bishop D.T., Spurr N.K. (1993) *Hum. Mol. Genet.* **2**, 1823–1828.
7. Goldgar D.E., Fields P., Lewis C.M., Tran T.D., Cannon-Albright L.A., Ward J.H., Swensen J., Skolnick M.H. (1994) *J. Natl Cancer Inst.* **86**, 200–209.
8. Smith S.A., Dicioccio R.A., Streuwing J.P., Easton D.F., Gallion H.H., Albertsen H., Mazoyer S., Johansson B., Steichen-Gersdorf E., Stratton M., Ford D., Marshall G., White R.L., Piver M.S., Ponder B.A.J. (1994) *Genes Chrom. Cancer* **10**, 71–74.
9. Luu-The V., Labrie C., Simard J., Lachance Y., Zhao H-F., Couet J., Leblanc G., Labrie F. (1990) *Mol. Endocrinol.* **4**, 268–275.
10. Feunteun J., Narod S.A., Lynch H.T., Watson P., Conway T., Lynch J., Parboosingh J., O'Connell P., White R., Lenoir G.M. (1993) *Am. J. Hum. Genet.* **52**, 736–742.
11. Cropp C.S., Nevanlinna H.A., Pyrhonen S., Stenman U-H., Salmikangas P., Albertsen H., White R., Callahan R. (1994) *Cancer Res.* **54**, 2548–2551.
12. Ford D., Easton D.F., Bishop D.T., Narod S.A., Goldgar D.E. (1994) *Lancet* **343**, 692–695.
13. Weber J.L., Kwitek A.E., May P.E., Wallace M.R., Collins F.S., Ledbetter D.H. (1990), *Nucleic Acid Res.* **18**, 4640.
14. Hall J.M., Friedman L., Guenther C., Lee M.K., Weber J.L., Black D.M., King M-C. (1992) *Am. J. Hum. Genet.* **50**, 1235–1242.
15. Weissenbach J., Gyapay G., Dio C., Vignal A., Morissette J., Millaseau P., Vaysseix G.A., Lathrop M.A. (1992) *Nature* **359**, 794–801.
16. Anderson L.A., Friedman L., Osborne-Lawrence S., Lynch E., Weissenbach J., Bowcock A., King M-C. (1993) *Genomics* **17**, 618–623.
17. Flejter W.L., Kukowska-Latallo J.F., Kioussis S., Chandrasekharappa S.C., King S.E., Chamberlain J.S. (1993) *Hum. Mol. Genet.* **2**, 1080.
18. Friedman L,S., Lynch E.D., King M.S. (1993) *Hum. Mol. Genet.* **2**, 821.
19. Normand T., Narod S.A., Labrie F., Simard J. (1993) *Hum. Mol. Genet.* **3**, 479–483.
20. Nakamura Y., Lathrop M., O'Connell P., Leppert M., Barker D., Wright E., Skolnick M., Kondoleon S., Litt M., Lalouel J-M, White R. (1988) *Genomics* **82**, 3343–3346.
21. Lathrop G. M., Lalouel J.-M., Julier C., Ott J. (1984) *Proc. Natl Acad. Sci. USA* **81**, 3443–3446.
22. Narod S. A. (1991) *Clin. Genet.* **39**, 125–131.

Lopes-Cendes I, Andermann E, Attig E, Cendes F, Bosch S, Wagner M, Gerstenbrand F, et al (1994) Confirmation of the *SCA-2* locus as an alternative locus for dominantly inherited spinocerebellar ataxias and refinement of the candidate region. Am J Hum Genet 54:774–781

Orozco-Diaz G, Fleites AN, Sagaz RC, Auburger G (1990) Autosomal dominant cerebellar ataxia: clinical analysis of 263 patients from a homogeneous population in Holguin, Cuba. Neurology 40:1369–1375

Orr HT, Chung M, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226

Ranum LPW, Schut LJ, Lundgren JK, Orr HT, Livingston DM (1994) Spinocerebellar ataxia type 5 in a family descended from the grandparents of President Lincoln maps to chromosome 11. Nat Genet 8:280–284

Sasaki H (1993) Linkage study of spinocerebellar ataxia and probable correlation for the loci to the disease phenotype. Rinsho Shinkeigaku 33:1285–1287

Stevanin G, Le Guern E, Ravisé N, Chneiweiss H, Dürr A, Cancel G, Vignal A, et al (1994) A third locus for autosomal dominant cerebellar ataxia type 1 maps to chromosome 14q24.3-qter: evidence for the existence of a fourth locus. Am J Hum Genet 54:11–20

Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, et al (1993) The gene for Machado-Joseph disease maps to human chromosome 14q. Nat Genet 4:300–304

Twells R, Gispert S, Orozco G, Brice A, Weber J, Heredero L, Schewfler K, et al (1993) Chromosomal assignment of a second locus for autosomal dominant cerebellar ataxia (SCA) to chromosome 12q23-24.1. Nat Genet 4:295–299

0002-9297/95/5701-0022$2.00

Am. J. Hum. Genet. 57:189, 1995

BRCA1 Mutations in Ashkenazi Jewish Women

To the Editor:

The gene for breast and ovarian cancer, *BRCA1*, has recently been cloned (Miki et al. 1994). To date, we have identified *BRCA1* mutations in 24 North American families. In a recent collaborative report, the most common *BRCA1* mutation found was a 2-bp deletion in exon 2 (Shattuck-Eidens et al. 1995). This mutation (185delAG) was present in 6 of our 24 breast-ovarian cancer families. All six families are of Ashkenazi Jewish origin. The families are not known to be related to each other, but haplotype analyses suggest that these six families have a common ancestor (Simard et al. 1994). Only 1 of the other 18 families with *BRCA1* mutations was known to be of Jewish origin or had founders with surnames suggestive of Ashkenazi heritage ($P < .0001$). In total, six of the seven Ashkenazi families were found to carry the 185delAG mutation. Three other Ashkenazi families with this mutation have been reported by Struewing et al. (1995; in this issue). If the majority of hereditary breast-ovary cancer families in any ethnic subgroup can be attributed to a small number of mutations, our efforts to provide DNA-based predictive testing will be greatly enhanced.

PATRICIA TONIN,[1] OLGA SEROVA,[2] GILBERT LENOIR,[2] HENRY LYNCH,[3] FRANCINE DUROCHER,[4] JACQUES SIMARD,[4] KENNETH MORGAN,[1] AND STEVEN NAROD[1]

[1]*Division of Medical Genetics, Department of Medicine, and Department of Human Genetics, McGill University, Montreal;* [2]*International Agency for Research on Cancer, Lyon;* [3]*Creighton University School of Medicine, Department of Preventative Medicine and Public Health, Omaha; and* [4]*Laboratory of Molecular Endocrinology, CHUL Research Centre and Laval University, Quebec*

Acknowledgments

This work was supported by grants from the Canadian Genetic Diseases Network (Federal NCE program), the National Cancer Institute of Canada, the National Breast Cancer Research Initiative, Endorecherche, and the Department of the Army (United States) grant DAMD17-94-J-4299. S.N. is a fellow of the Fonds de la Recherche en Sante du Québec. J.S. is a scholar the Medical Research Council of Canada (MRC). F.D. is a recipient of a studentship from the MRC.

References

Miki Y, Swenson J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavitgan S, Lui Q, et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science 266:66–71

Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, Hoskins K, et al (1995) A collaborative survey of 80 mutations in the *BRCA1* breast and ovarian cancer susceptibility gene. JAMA 273:535–541

Simard J, Tonin P, Durocher F, Morgan K, Rommens J, Gingras S, Samson C, et al (1994) Common origins of *BRCA1* mutations in Canadian breast and ovarian cancer families. Nat Genet 8:392–398

Struewing JP, Brody LC, Erdos MR, Kase RG, Giambarresi TR, Smith SA, Collins FS, et al (1995) Detection of eight *BRCA1* mutations in 10 breast/ovarian cancer families, including 1 family with male breast cancer. Am J Hum Genet 57:1–7 (in this issue)

0002-9297/95/5701-0023$2.00